27èmes rencontres HélioSPIR

Toulouse, 23-24 juin 2026

Résumés des communications

Association HélioSPIR
Réseau de spectroscopie proche infrarouge
www.heliospir.net

HélioSPIR est l'association francophone dédiée à la spectrométrie dans le proche infrarouge.

HélioSPIR a vocation à fédérer les scientifiques et les utilisateurs de la technologie SPIR au sein d'un réseau et à promouvoir l'utilisation de la spectroscopie proche infrarouge. Fondée en 2004 autour de la communauté scientifique d'Agropolis à Montpellier, l'association dépasse maintenant les contours de la région Occitanie et de l'hexagone. C'est un pôle de compétences à dimension internationale dans le domaine de la spectroscopie proche infrarouge.

HélioSPIR organise chaque année une ou deux sessions de rencontres scientifiques. C'est un moment privilégié d'échanges autour de diverses thématiques autour de la spectroscopie proche infrarouge et de découverte des derniers travaux de la communauté. C'est également l'occasion de découvrir ou redécouvrir les équipements de spectroscopie et d'imagerie hyperspectrales des principaux fabricants du secteur.

Comment citer ce document

HélioSPIR, 2026. Résumés des communications présentées aux 27èmes rencontres HélioSPIR, Toulouse (France), 23-24 juin 2026. D. Bastianelli, G. Chaix, Eds. Association HélioSPIR, Montpellier (France), 55p. DOI : 10.19182/agritrop/00248

Comment citer un résumé particulier

Auteur 1, Auteur 2… Auteur n, 2026. Titre du résumé. *In :* HélioSPIR, 2026. Résumés des communications présentées aux 27èmes rencontres HélioSPIR, Toulouse (France), 23-24 juin 2026. D. Bastianelli, G. Chaix, Eds. (DOI : 10.19182/agritrop/00248), Association HélioSPIR, Montpellier (France). Numéros de page.

Communications orales

Pages

Mardi 23 juin 2026

B. Aernouts et al. An update on our NIR milk quality research, for instance on the use of DOP and slope and bias correction to achieve accurate predictions over years 5

M. Ryckewaert. REP-ASCA : Analyse de données de spectroscopie VIS-NIR issues de plans d'expériences en agronomie 10

A. Maléchaux et al. Surmonter les limites de la spectroscopie proche infrarouge embarquée : Synergie entre pré-traitement, actualisation annuelle et Machine Learning appliquée au phénotypage *in situ* du maïs 13

S. Treguier et al. Identification de bactéries lactiques sur gélose nutritive par spectroscopie UV-Vis-NIR 15

V Larat. Substitution de méthodes analytiques par des prédictions NIRS en alimentation animale 16

A. Bourely. 40 ans d'innovations pour la Planète : robots, SPIR, IA,... 17

M. Metz et al. Apprentissage en contexte : une solution pour combiner différentes bases de données spectrales ? 18

F. Abdelghafour et al. Chimiométrie profonde & détection d'anomalie : apprentissage autosupervisé pour l'authentification de produit alimentaire 19

D. Tanzilli. Attention-based aggregation of preprocessing for spectral deep learning 21

J. Brustel. Didactical use of PLS, Local PLS, SVM and ANN regressions: stevia and soybean 23

J.M. Roger. DROP-DA: A discriminant model that avoids overfitting 24

Mercredi 24 juin 2026

S. Cuq et al. Diagnostic nutritionnel en micro et macronutriments des limbes, pétioles et baies de raisin de vitis vinifera par l'utilisation du proche-infrarouge et d'outils de chimiométrie 25

S. Kamgaing et al. Transfert d'étalonnages NIR du diagnostic foliaire du palmier à huile pour le déploiement d'un réseau de micro-spectromètres au champ 26

M. Rosa et al. Comparaison des approches de régression PLS et d'apprentissage automatique pour la caractérisation chimique de l'huile de palme rouge par spectroscopie proche infrarouge 28

L.M.B. Tavares et al. Near-Infrared Spectroscopy for wood species discrimination and cardanol binder detection in biomass pellets 30

N. Cardoso Pereira et al. Impact of water stress on wood formation for Eucalyptus grandis as revealed by NIR spectroscopy 32

A. Arnal et al. Distinction de maladies de conservation par NIRS sur pommes 34

M. Hildebrandt et al. Appui de la Spectroscopie Proche Infrarouge pour l'amélioration de l'identification et la discrimination des espèces d'arbres sur pied en forêt amazonienne 36

I. Tumoine et al. Comment réduire la dépendance des réseaux de neurones aux valeurs de référence ? Les auto-encodeurs masqués, un cas appliqué d'apprentissage auto-supervisé 39

T. Randriambinintsoa et al. Développement d'un réseau de micro-spectromètres NIR pour la discrimination d'espèces forestières malgaches 41

M.A. Antar et al. Étude de l'adultération des épices en utilisant la spectroscopie visible et proche infrarouge combinée à la chimiométrie et l'apprentissage automatique 45

Posters

Pages

A. Bled et al. Mesures *in situ* par SPIR de la qualité de chênes pédonculés abroutis ou non par le chevreuil 46

E. Legros et al. Early prediction of tuber yield in yam (Dioscorea spp.) using NIR spectra from pre-planting tubers and mature leaves 48

M. Lesnoff. Jchemo: Chemometrics and machine learning on high-dimensional data with Julia 51

M. Ribes et al. PRO-PIX / ONE-PIX : une approche d'imagerie hyperspectrale mono-pixel pour la spectroscopie appliquée et la production d'indicateurs embarquée 53

Communications orales

An update on our NIR milk quality research: methods to achieve accurate predictions over years

Ben Aernouts[1*], Leandro P. da Silva[2a], José A. Diaz-Olivares[1], Xinyue Fu[1], Arnout van Nuenen[1] and Javier E. L. Villa[2]

[1]KU Leuven, Department of Biosystems, Division of Animal and Human Health Engineering, Campus Geel; Kleinhoefstraat 4, 2440 Geel, Belgium
[2]University of Campinas (UNICAMP), Institute of Chemistry, Group of Chemometrics and Applied Spectroscopy; Campinas 13081-970, SP, Brazil

** Corresponding and presenting author: ben.aernouts@kuleuven.be*
[a] Main author of this work

***Keywords** : NIR milk quality analysis, Multivariate calibration, Instrumental drift, Domain shift.*

Abstract

Near-infrared (NIR) spectroscopy combined with Partial Least Squares Regression (PLSR) enables rapid, non-destructive prediction of milk composition in automated milking systems. Under practical farm conditions, seasonal, biological, and management-related variability introduce domain shifts between calibration and prediction datasets, leading to performance degradation over time.

This study evaluates two calibration transfer strategies, CORrelation ALignment (CORAL) and Dynamic Orthogonal Projection (DOP), to improve the robustness of NIR milk composition models across temporally distinct milk populations. Milk samples (N = 6,100) were collected on-farm and divided into six successive temporal datasets. PLSR models were calibrated on the first dataset and transferred to subsequent datasets. Performance was evaluated using the root mean squared error of prediction (RMSEP), standard error of prediction (SEP), bias, and slope for fat, protein, and lactose. Posterior bias and slope correction (BSC) were assessed both independently and in combination with CORAL and DOP.

Direct model transfer increased SEP and introduced systematic deviations, confirming the presence of domain shift. BSC alone yielded limited improvement. CORAL reduced prediction error through unsupervised covariance alignment, whereas DOP achieved stronger correction by removing structured spectral drift using representative transfer samples. The combination of DOP and BSC provided the most consistent performance recovery. These results emphasise the importance of calibration transfer strategies for reliable long-term on-farm deployment.

Introduction

Accurate quantification of milk composition is essential for quality control, economic optimisation, and monitoring animal health and welfare in modern dairy systems. Continuous monitoring of fat, protein, and lactose supports informed decision-making at the farm level and enables real-time control in automated milking systems. Near-infrared (NIR) spectroscopy (1000–1700 nm), combined with Partial Least Squares

Regression (PLSR), provides a rapid and non-destructive method for on-farm prediction of major milk components [1].

However, under real farm conditions, spectral variability arises from seasonal changes, herd composition, feeding regimes, and management practices. These factors introduce discrepancies between calibration and prediction datasets, resulting in domain shift and reduced predictive performance when models are applied to spectra of new milk samples [2,3].

Calibration transfer techniques aim to mitigate these discrepancies between source (calibration) and target (prediction) domains. CORrelation ALignment (CORAL) is an unsupervised approach that reduces domain shift by aligning the covariance structure of source and target data in the original feature space [4]. In contrast, Dynamic Orthogonal Projection (DOP) is a supervised method that removes structured spectral deviations using representative transfer samples [5].

Although both approaches address domain shift, they are based on different principles, and their relative effectiveness for transferring NIR milk composition models across successive farm datasets remains insufficiently explored. Therefore, this study evaluates and compares CORAL and DOP for calibration transfer of NIR-based PLSR models across temporally distinct milk populations. Model performance is assessed by prediction accuracy for fat, protein, and lactose to support robust spectroscopic deployment in automated milking systems.

Materials and methods

Milk samples were collected at the dairy farm Hooibeekhoeve (Geel, Belgium) using an automatic milking robot (VMS™ Classic, DeLaval, Tumba, Sweden). An improved NIR spectroscopic sensor, previously described in [6], was integrated into the system and connected to a DeLaval Herd Navigator™ sampler to ensure representative full-milking sampling [3]. Approximately 50 mL of milk was directed to the spectroscopic unit during each milking, while an aliquot (~30 mL) was collected for laboratory reference analysis.

Samples were temperature-stabilised at 38 °C and analysed in a 2 mm flow-through quartz cuvette. Transmittance spectra (950–1700 nm) were acquired using a halogen light source between July 2022 and January 2024. Each measurement cycle included sample, dark, and reference scans to correct instrumental drift and detector sensitivity. The data were grouped into six temporal milk populations: G1 (N = 911), G2 (N = 303), G3 (N = 783), G4 (N = 1,407), G5 (N = 1,560) and G6 (N = 218). The definition of the temporal groups was based on an exploratory analysis of both the sample transmittance spectra and the white reference spectra, aimed at identifying structural changes throughout the monitored period. G6 was left completely untouched and reserved exclusively for independent external validation. Each sample consisted of paired spectra and reference values for fat, protein, and lactose.

Spectral regions 1405–1500 nm (water absorption) and 1675–1700 nm (low detector sensitivity) were excluded. Preprocessing included Standard Normal Variate (SNV) followed by a Savitzky–Golay second derivative (second-order polynomial, 11-point window). Within G1, samples were split into calibration and validation sets using the SPXY algorithm [7]. PLSR models were trained on the calibration subset of G1 with an optimised number of latent variables, and subsequently applied to G2–G6.

Domain shift was addressed using CORAL and DOP. CORAL aligned the covariance structure of target spectra to that of the G1 calibration set through a whitening–recolouring transformation, without using target reference values [4].

DOP was implemented as a supervised transfer method. 80 representative transfer samples ("spike samples") were selected using SPXY, and an interference subspace was estimated from deviations between source and target spectra. Target spectra were then orthogonally projected to remove structured variation before prediction [5]. Model parameters were optimised via grid search.

Posterior bias and slope correction (BSC) was applied after CORAL and DOP using the subset of spike samples. BSC was also evaluated independently on the original predictions.

(next page)

Table 1 : Root mean square error of prediction (RMSEP), standard error of prediction (SEP), bias and slope for the original model (BASE), after correlation alignment (CORAL), dynamic orthogonal projection (DOP), bias and slope correction (BSC), CORAL+BSC and DOP+BSC for prediction of milk fat, protein and lactose for the test set of group 1 (G1), for group 2 (G2), group 3 (G3), group 4 (G4), and group 5 (G5), and for the independent external validation on group 6 (G6).

Milk component	Group	RMSEP						SEP						Bias						Slope					
		BASE	CORAL	DOP	BSC	C+BSC	D+BSC	BASE	CORAL	DOP	BSC	C+BSC	D+BSC	BASE	CORAL	DOP	BSC	C+BSC	D+BSC	BASE	CORAL	DOP	BSC	C+BSC	D+BSC
Fat	G1	0.065	0.065	0.065	0.065	0.065	0.065	0.064	0.064	0.064	0.064	0.064	0.064	-0.012	-0.012	-0.012	-0.012	-0.012	-0.012	0.994	0.994	0.994	0.994	0.994	0.994
	G2	0.098	0.271	0.425	0.089	0.104	0.092	0.09	0.119	0.09	0.088	0.102	0.09	0.038	0.244	-0.415	-0.016	-0.02	-0.023	1.02	1.06	0.975	0.972	0.965	0.989
	G3	0.186	0.268	0.081	0.106	0.118	0.072	0.105	0.125	0.07	0.104	0.116	0.071	0.154	0.238	-0.04	0.018	0.019	0.012	1.015	1.041	0.986	0.955	0.952	0.977
	G4	0.368	0.406	0.115	0.082	0.083	0.068	0.083	0.103	0.068	0.082	0.083	0.068	0.359	0.393	-0.093	-0.005	-0.005	-0.005	0.963	1.077	0.993	0.996	0.995	0.994
	G5	0.247	0.269	0.101	0.103	0.103	0.101	0.106	0.105	0.1	0.102	0.103	0.101	0.222	-0.248	0.009	-0.011	-0.012	0.003	0.947	0.958	0.989	0.978	0.977	0.997
	G6	24.43			0.53		0.12	2.83			0.53		0.12	24.27			-0.04		0	-2.13			0.65		0.97
Protein	G1	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	0.927	0.927	0.927	0.927	0.927	0.927
	G2	0.188	0.123	0.36	0.109	0.113	0.111	0.118	0.119	0.116	0.108	0.111	0.11	0.147	-0.029	-0.341	-0.021	-0.023	-0.02	1.034	1.017	1.01	0.853	0.85	0.833
	G3	0.204	0.234	0.729	0.148	0.15	0.117	0.203	0.2	0.133	0.148	0.15	0.117	-0.021	-0.121	0.717	-0.009	-0.009	-0.008	1.12	1.089	1.024	0.722	0.715	0.848
	G4	0.719	0.12	0.97	0.106	0.108	0.097	0.108	0.11	0.1	0.106	0.108	0.097	0.711	-0.049	0.965	-0.007	-0.007	0.003	0.895	0.899	0.968	0.835	0.837	0.923
	G5	0.657	0.315	0.282	0.155	0.154	0.097	0.114	0.115	0.103	0.152	0.151	0.096	0.647	-0.293	0.263	-0.03	-0.03	-0.014	0.863	0.861	1.013	0.524	0.529	0.89
	G6	21.49			0.27		0.1	1.97			0.27		0.1	-21.4			0.05		0.01	3.77			0.33		1.01
Lactose	G1	0.058	0.058	0.058	0.058	0.058	0.058	0.057	0.057	0.057	0.057	0.057	0.057	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.946	0.946	0.946	0.946	0.946	0.946
	G2	1.192	0.212	0.356	0.102	0.1	0.067	0.192	0.182	0.076	0.102	0.101	0.067	-1.177	-0.11	-0.347	0.001	0.001	-0.008	1.346	1.312	1.074	0.523	0.533	0.94
	G3	0.992	0.152	0.093	0.088	0.088	0.072	0.11	0.107	0.078	0.088	0.088	0.072	-0.986	-0.108	0.05	0.003	0.003	0.003	0.98	0.971	0.952	0.549	0.551	0.836
	G4	1.388	0.146	1.088	0.088	0.092	0.074	0.088	0.092	0.07	0.087	0.091	0.072	-1.385	-0.113	-1.086	0.01	0.008	0.015	0.87	0.864	0.902	0.866	0.856	0.945
	G5	1.087	0.13	0.675	0.122	0.122	0.091	0.111	0.111	0.1	0.121	0.121	0.091	-1.081	-0.069	-0.667	0.018	0.018	0.005	0.943	0.937	0.924	0.218	0.221	0.795
	G6	56.53			0.19		0.11	8.01			0.19		0.1	55.97			0.05		0.05	5.22			0.02		0.68

Results and discussion

Direct application of G1-calibrated models (BASE) to G2–G6 increased SEP and introduced systematic bias and slope deviations, confirming domain shift. The magnitude of degradation varied across analytes, indicating differing sensitivity to spectral drift.

Applying BSC alone reduced intercept and slope errors but only partially improved SEP, suggesting that domain shift is not purely linear. CORAL decreased prediction error by correcting global covariance differences, though residual bias remained in some cases.

DOP provided a stronger recovery when spectral differences were structured, as it explicitly removed interference components. Notably, effective correction was achieved using a limited number of 80 transfer samples. The additional improvement observed when combining DOP with BSC indicates that small residual linear effects remain after spectral correction.

Conclusions

Domain shift significantly impacts NIR milk composition predictions across temporally distinct milk populations. While linear recalibration alone offers limited improvement, CORAL and DOP enhance model robustness. DOP provides the most effective correction when transfer samples are available, supporting the integration of calibration transfer techniques for reliable on-farm spectroscopic applications.

Acknowledgments

The authors would like financial support from the São Paulo Research Foundation (FAPESP—grants 2021/05679-8 and 2024/09412-4) and the National Council for Scientific and Technological Development (CNPq—grants 406607/2022-2 and 141280/2023-9).

References

[1] B., Aernouts, et al. Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance? *Journal of Dairy Science*, Volume 94(11), 2011, 5315–5329, https://doi.org/10.3168/jds.2011-4354.

[2] A. van Nuenen, et al. On-farm NIR sensor for milk analysis: Exploitation of bias monitoring and bias correction. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, Volume 320, 2024, 124544, https://doi.org/10.1016/j.saa.2024.124544.

[3] J. A., Diaz-Olivares, et al. Temperature correction of near-infrared spectra of raw milk. *Chemometrics and Intelligent Laboratory Systems*, Volume 255, 2020, 105251, https://doi.org/10.1016/j.chemolab.2024.105251.

[4] B. Sun, et al. Correlation Alignment for Unsupervised Domain Adaptation, In: Csurka, G. (eds) *Domain Adaptation in Computer Vision Applications. Advances in Computer Vision and Pattern Recognition*. Springer, Cham, 2017, pp. 153–171. https://doi.org/10.1007/978-3-319-58347-1_8

[5] M., Zeaiter, et al. Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. *Chemometrics and Intelligent Laboratory Systems*, Volume 80(2), 2006, 227–235, https://doi.org/10.1016/j.chemolab.2005.06.011.

[6] J. A., Diaz-Olivares, et al. Online milk composition analysis with an on-farm near-infrared sensor. *Computers and Electronics in Agriculture*, Volume 178, 2020, 105734, https://doi.org/10.1016/j.compag.2020.105734.

[7] R., Galvao, et al. A method for calibration and validation subset partitioning. *Talanta*, Volume 67, 2005, 736–740, https://doi.org/10.1016/j.talanta.2005.03.025.

REP-ASCA : Analyse de données de spectroscopie VIS-NIR issues de plans d'expériences en agronomie

Maxime RYCKEWAERT[1,2]

[1] *CIRAD - UMR AGAP Institut, Montpellier, France*
[2] *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*

Email : maxime.ryckewaert@cirad.fr

***Mots-clefs** : REP-ASCA, Analyse de variance, données spectrales*

Introduction

L'analyse de variance appliquée aux données spectrales se heurte à des contraintes structurelles : ratio p >> n, forte colinéarité entre longueurs d'onde, non-normalité des distributions. La méthode ASCA [1] résout ces problèmes en décomposant la matrice spectrale X selon les facteurs du plan d'expérience et en testant leur significativité par permutation, sans hypothèse distributionnelle. Pour chaque facteur significatif, une analyse en composantes simultanées (SCA) fournit scores et composantes caractérisant la séparation entre niveaux et les régions spectrales impliquées.

En spectroscopie, les mesures répétées d'un même échantillon présentent des différences systématiques d'origine physique : décalages de ligne de base, effets multiplicatifs liés à la taille des particules ou à la rugosité de surface, fluctuations d'illumination en conditions terrain. Dans le cadre ASCA, cette erreur de répétabilité gonfle la variance résiduelle SSQ(E) et dilue les variances factorielles, pouvant rendre non détectables des effets biologiquement réels. Les prétraitements standards (SNV, MSC, dérivées) corrigent partiellement ces effets mais de façon empirique, sans garantie que l'information chimique pertinente est préservée.

REP-ASCA [2] modélise explicitement le sous-espace spectral associé à l'erreur de répétabilité à partir de mesures répétées dédiées, puis projette orthogonalement les données expérimentales hors de ce sous-espace avant l'ASCA.

Matériel et méthodes

REP-ASCA :

Un jeu additionnel $\mathbf{X_s}$ de mesures répétées sur un sous-ensemble représentatif d'échantillons est acquis. Les spectres répétés sont centrés par échantillon et empilés dans une matrice $\mathbf{W}$ qui ne contient que la variabilité de répétabilité. Une ACP sur $\mathbf{W}$ fournit les k premières composantes (matrice $\mathbf{D}$) engendrant le sous-espace d'erreur. Les données expérimentales sont ensuite corrigées par projection orthogonale : $\mathbf{X}_\perp = \mathbf{X}(\mathbf{I} - \mathbf{D}\mathbf{D}^\mathbf{T})$. ASCA est appliquée à $\mathbf{X}_\perp$. Le paramètre k est sélectionné en maximisant les SSQ factorielles.

Cas d'études :

Trois cas d'étude sont présentés. Le premier [2] porte sur la torréfaction de grains de café (NIR 1000-2500 nm) selon un plan croisant l'espèce (Arabica, Robusta) et le temps de torréfaction (0, 25, 50, 75 min), avec 7 mesures par échantillon (n = 56 spectres) et un jeu W de 24 spectres répétés. Le deuxième [3] concerne du maïs sous stress hydrique mesuré en conditions terrain (VIS-NIR 310-1100 nm) selon un plan 10 génotypes x 2 traitements (irrigué/non irrigué), avec 12 spectres par micro-parcelle (n = 480 spectres), analysé à la date de stress maximal (30/07/2018, Nérac). Le troisième [4] utilise de l'imagerie

hyperspectrale VIS-NIR (407-997 nm, 186 bandes) sur Arabidopsis thaliana selon un plan traitement (irrigué vs stress hydrique) x date (11 jours d'acquisition sur 2 mois), avec un pipeline incluant normalisation VSN, ACP, segmentation U-Net et k-means avant REP-ASCA.

Résultats et discussion

Sur le jeu café, l'erreur de répétabilité représente 82 % de la variance totale dans l'ASCA standard, rendant tous les facteurs non significatifs ($p > 0,05$). L'ACP sur $\mathbf{W}$ identifie trois composantes d'origine physique ; décalage constant (91,6 %), pente (6,5 %) et courbure (1,1 %) ; dont les loadings ont la forme de lignes de base et non de pics chimiques, ce qui valide leur suppression. Après projection orthogonale avec $k = 5$, les trois facteurs deviennent significatifs et les loadings de l'effet espèce révèlent des bandes interprétables : teneur en eau (1930 nm), acides gras (1728 et 1763 nm), caféine et glucides (1212 nm), acides chlorogéniques (2070-2238 nm). Ce cas illustre la situation la plus défavorable, où le bruit physique domine à tel point que l'ASCA standard est totalement inopérante sans correction préalable.

Sur le jeu maïs, la variance résiduelle atteint 65 % avant correction et le facteur traitement n'explique que 4,5 % de la variance. Les deux composantes de $\mathbf{W}$ correspondent à un décalage global (lié à l'angle de vue) et à une opposition VIS/NIR avec un pic négatif à 709 nm interprétable comme une variabilité de position du red-edge liée à l'angle foliaire au sein de la micro-parcelle, effet difficilement corrigeable par les prétraitements standards. Avec $k = 2$, tous les facteurs deviennent significatifs. Les loadings de l'effet traitement impliquent le red-edge (739 nm), la bande eau (970 nm) et la région UV (300-420 nm, accumulation de flavonoïdes sous stress). Le terme d'interaction génotype-par-traitement corrèle à $R = 0,81$ avec la perte de rendement en environnement stressé et à $R = 0,60$ lors du transfert vers un environnement de faible stress, suggérant que REP-ASCA extrait une information génotypique stable pertinente pour la « phénotypisation » haut-débit en conditions terrain.

Sur le jeu Arabidopsis, REP-ASCA détecte un effet traitement significatif dès 3 jours après induction du stress hydrique, avant l'apparition de symptômes visuels, avec des loadings impliquant la région des anthocyanines (550 nm) et le red-edge (670-720 nm). Ce résultat confirme que la méthode s'étend naturellement à l'imagerie hyperspectrale, où l'erreur d'illumination entre sessions joue un rôle analogue à l'erreur instrumentale classique, et qu'elle est sensible à des modifications biochimiques précoces.

Les trois cas d'étude convergent vers un même constat : l'interprétabilité des loadings de $\mathbf{W}$ constitue un avantage opérationnel central par rapport aux prétraitements empiriques, car elle permet de vérifier explicitement que les composantes supprimées sont d'origine physique. La sélection de k reste néanmoins dépendante de l'expertise de l'utilisateur ; le développement d'un critère automatique basé sur la forme spectrale des loadings ou sur la colinéarité entre sous-espace de W et matrices factorielles représente une perspective directe. REP-ASCA s'inscrit dans la lignée de l'orthogonalisation par paramètres externes (EPO) tout en intégrant la structure du plan d'expérience et les tests de permutation.

Le code Python est disponible à https://github.com/RYCKEWAERT/REP-ASCA-Python .

Remerciements

Federico Marini et Jean-Michel Roger, Daphné Héran, Ryad Bendoula, Puneet Mishra, Fabienne Henriot, Alexia Gobrecht et Nathalie Gorretta

References

[1] A. K. Smilde, J. J. Jansen, H. C. J. Hoefsloot, R.-J. A. N. Lamers, J. Van Der Greef, et M. E. Timmerman, « ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data », *Bioinformatics*, vol. 21, n° 13, p. 3043-3048, juill. 2005, doi: 10.1093/bioinformatics/bti476.

[2] M. Ryckewaert, N. Gorretta, F. Henriot, F. Marini, et J.-M. Roger, « Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample », *Anal. Chim. Acta*, vol. 1101, p. 23-31, 2020.

[3] M. Ryckewaert *et al.*, « Potential of high-spectral resolution for field phenotyping in plant breeding: Application to maize under water stress », *Comput. Electron. Agric.*, vol. 189, p. 106385, 2021.

[4] P. Mishra *et al.*, « A generic workflow combining deep learning and chemometrics for processing close-range spectral images to detect drought stress in Arabidopsis thaliana to support digital phenotyping », *Chemom. Intell. Lab. Syst.*, vol. 216, p. 104373, 2021.

Surmonter les limites de la spectroscopie proche infrarouge embarquée : Synergie entre pré-traitement, actualisation annuelle et Machine Learning appliquée au phénotypage *in situ* du maïs.

Astrid Maléchaux[1], Magali Roussel[2], Jordane Poulain[1], Milagros Garcia[2], Sylvie Roussel[1]

[1] *Ondalys, 8 Av. de l'Europe, 34830 Clapiers, France*
[2] *LIDEA Seeds, 6 Chemin de Panedautes, 31700 Mondonville, France*

Email : amalechaux@ondalys.fr, magali.roussel@lidea-seeds.com

***Mots-clefs** : NIRS embarqué, régression SVM, orthogonalisation, recalage annuel*

Introduction

Dans le cadre du programme de sélection variétale du maïs (réseau expérimental LIDEA), l'évaluation de critères biochimiques constitue une étape clé du processus de sélection. Si le phénotypage traditionnel repose sur une logistique lourde (échantillonnage in situ, préparation en station et centralisation pour analyse par spectroscopie proche infrarouge (NIRS) en laboratoire), l'intégration de spectromètres NIRS directement embarqués sur les moissonneuses-batteuses offre une alternative stratégique pour un criblage exhaustif à coût réduit.

Cependant, l'évaluation *in situ* de caractères biochimiques simples ou complexes — résultant de l'interaction de plusieurs fractions biochimiques — au sein d'une matrice hétérogène se heurte à de multiples sources de variabilité : hétérogénéité physique des échantillons, variabilité inter-instruments et variations environnementales liées à la récolte. Face à ce bruit de fond, les approches chimiométriques classiques (prétraitements standards et régression PLS) s'avèrent insuffisantes pour fournir des prédictions précises et robustes. Pour lever ce verrou technologique, cette étude déploie une méthodologie combinant prétraitements spectraux avancés, modélisation non linéaire et enrichissement dynamique. Cette synergie permet d'évaluer ces phénotypes complexes directement sur le terrain, en exploitant les données NIRS de laboratoire comme méthode de référence.

Matériel et méthodes

L'étude s'appuie sur une base de données de plus de 56 000 spectres collectés in situ entre 2017 et 2022, en différents lieux et avec différents spectromètres embarqués sur plusieurs moissonneuses-batteuses. Après l'élimination des données atypiques, une chaîne séquentielle optimisée de prétraitements a été développée pour isoler le signal analytique : dérivée seconde de Savitzky-Golay, correction SNV et orthogonalisation EROS (Error Removal by Orthogonal Substraction) pour soustraire spécifiquement la variance liée aux instruments et aux conditions environnementales.

Une approche prédictive comparative a ensuite été menée sur un jeu de test indépendant afin d'évaluer les performances de la régression PLS face à des modèles non linéaires à vecteurs de support (SVM). Le modèle PLS a été optimisé par une stratégie de validation-croisée prenant en compte les groupes combinant année, machine, spectromètre et lieu. Pour des raisons de temps de calcul, les paramètres du modèle SVM ont été optimisés sur un jeu de données réduit sélectionné par la méthode Kennard-Stone de façon à être représentatif du jeu d'étalonnage complet.

Dans un premier temps, les performances des modèles ont été comparées sur un jeu de validation interne indépendant constitué de 310 échantillons au total, soit 10 échantillons de chaque combinaison d'année, machine, spectromètre et lieu pour les années présentes dans le jeu d'étalonnage. La robustesse des

modèles a ensuite été testée sur les mesures d'une nouvelle année non représentée dans le jeu d'étalonnage.

Enfin, une stratégie d'actualisation dynamique par recalage annuel a été implémentée. Elle repose sur l'intégration parcimonieuse de données de l'année en cours dans le jeu d'apprentissage, via un sous-échantillonnage ciblé (plafonné à 25 % des flux, maximum 3000 échantillons) fondé sur un tirage aléatoire stratifié de 45 échantillons par strate combinant machine, spectromètre, lieu et date de mesure. Ces échantillons sont utilisés pour mettre à jour les calculs d'orthogonalisation, puis les paramètres du modèle de régression.

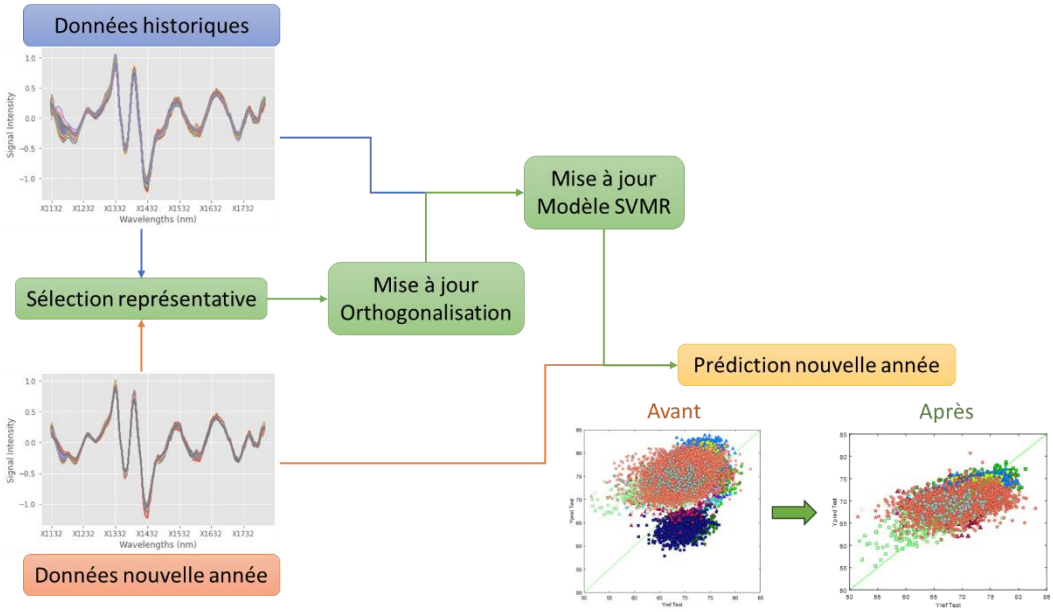


Figure 1 : Schéma de la stratégie d'actualisation annuelle du modèle de régression SVM

Résultats et discussion

Les résultats mettent en évidence l'incapacité des approches linéaires (PLS) à capter la complexité du signal environnemental et matriciel, avec un RMSEP de 2.6% sur le jeu de validation interne contre 7.9% lors de l'extrapolation à une nouvelle année. Le modèle SVM offre de meilleures performances en validation interne (RMSEP de 2.4%), mais subit néanmoins une dérive significative lors de l'extrapolation stricte à une nouvelle campagne de récolte (RMSEP de 5.8%), illustrant le poids de l'effet annuel.

Pour y remédier, la stratégie d'actualisation par recalage annuel a été implémentée sur le modèle SVM. L'intégration de cet ancrage in situ a permis de corriger substantiellement les prédictions, réduisant l'erreur d'extrapolation à un seuil acceptable de 3.1%, proche de l'incertitude intrinsèque de la méthode de référence. La robustesse de ce pipeline (Prétraitements + Recalage + SVM) a été validée avec succès en routine lors des campagnes 2023, 2024 et 2025, affichant des performances stables dans le temps.

Conclusion

Ce pipeline analytique confirme le potentiel du NIRS embarqué pour le phénotypage haut débit de caractères complexes. Sur le plan opérationnel, il assure un criblage précoce et robuste, permettant d'écarter les variétés peu performantes dès les premières phases de sélection. Transposable à d'autres caractères d'intérêt agronomiques, cette stratégie réduit significativement l'empreinte logistique et financière des programmes d'amélioration variétale.

Identification de bactéries lactiques sur gélose nutritive par spectroscopie UV-Vis-NIR

Sylvain Tréguier [1], Christel Couderc [1], Marjorie Audonnet [2], Hélène Tormo [1], Marie-Line Daveran-Mingot [2], Didier Kleiber [1], Cécile Levasseur-Garcia [3]

[1] Ecole d'Ingénieurs de Purpan, Université de Toulouse, INPT, Toulouse, France
[2] TBI, CNRS, INRA, INSA, Université de Toulouse, Toulouse, France
[3] Laboratoire de Chimie Agro-industrielle (LCA), Université de Toulouse, INRA, INPT, INP-PURPAN, Toulouse, France
Email : s.treguier@terresinovia.fr

***Mots-clefs** : bactéries lactiques, gélose, UV-Visible, SPIR, criblage*

Les bactéries lactiques jouent un rôle important dans de nombreux processus biologiques comme la fermentation, et suscitent donc un intérêt particulier dans l'industrie agroalimentaire. Certaines espèces bactériennes présentes dans le lait cru telles que *Lactococcus lactis* sont utilisées comme ferments dans la fabrication de produits laitiers et contribuent à leurs propriétés organoleptiques [1-3]. Les techniques actuelles d'identification des bactéries nécessitent une préparation spécifique des échantillons et la destruction de ces derniers au cours de l'analyse. Dans ce travail, nous proposons une méthode rapide et non destructive d'identification de bactéries lactiques à l'échelle du genre et de l'espèce. 84 souches ont été isolées de laits de chèvre crus par l'Ecole d'Ingénieurs de Purpan et caractérisées par le TBI (Toulouse Biotechnology Institute, Bio & Chemical Engineering, France). L'ensemble de ces souches, ainsi que 142 souches issues de collections, ont été inoculées sur des boîtes de Pétri contenant de la gélose nutritive. Les cultures bactériennes obtenues ont été analysées en réflectance par un spectromètre combinant les gammes optiques de l'UV-visible et du proche infrarouge.

Des modèles de classification distincts ont été développés pour le criblage du genre et de l'espèce. Dans les deux cas, les modèles de classification par ANN précédés d'une réduction de la dimensionnalité par ACP ont donné de meilleurs résultats en validation externe que les modèles d'ANN sans réduction de dimensions et les modèles de PLS-DA. Différentes approches de prétraitement ont été explorées pour éliminer ou réduire la signature spectrale de la gélose sur les acquisitions. Les meilleures performances pour l'identification du genre ont été obtenues en appliquant une EMSC (Extended Multiplicative Signal Correction) avec un spectre de gélose pure, tandis que les spectres bruts se sont avérés plus adaptés pour l'identification de l'espèce. Les prétraitements utilisés lors de cette étude pourraient engendrer une perte d'informations spectrales associées à l'espèce sur nos échantillons.

Les résultats de cette approche sont encourageants et témoignent de la capacité de la spectroscopie UV-visible-NIR à identifier le genre et l'espèce de souches bactériennes inconnues sur milieu gélosé. Un travail d'optimisation et une consolidation du jeu de données seront toutefois nécessaires pour pouvoir s'approcher des performances des techniques de diagnostic conventionnelles.

Références

[1] Caridi, A., Micari, P., Caparra, P., Cufari, A., & Sarullo, V. (2003). Ripening and seasonal changes in microbial groups and in physico-chemical properties of the ewes' cheese Pecorino del Poro. International Dairy Journal, 13, 191-200. doi:10.1016/S0958-6946(02)00157-7

[2] Welman, A. D., & Maddox, I. S. (2003). Exopolysaccharides from lactic acid bacteria: perspectives and challenges. Trends in Biotechnology, 21(6), 269-274. doi:10.1016/S0167-7799(03)00107-0

[3] Garnier, L., Mounier, J., Lê, S., Pawtowski, A., Pinon, N., Camier, B., Chatel, M., Garric, G., Thierry, A., Coton, E., Valence, F. (2019). Development of antifungal ingredients for dairy products: From in vitro screening to pilot scale application. Food Microbiology, 81, 97-107. doi:10.1016/j.fm.2018.11.003

Substitution de méthodes analytiques par des prédictions NIRS en alimentation animale

Vincent Larat [1]

[1] *Adisseo, CERN, 03600 Malicorne*

Email : vincent.larat@adisseo.com

***Mots-clefs** : Alimentation animale, Etudes in-vivo, Spectroscopie proche infrarouge*

La spectroscopie proche-infrarouge est utilisée depuis de nombreuses décennies dans le domaine de l'alimentation animale. Dans ce contexte, Adisseo a développé un service dédié (PNE, Precise Nutrition Evaluation) destiné à ses clients, leur permettant d'évaluer, à partir du spectre proche infrarouge, les caractéristiques nutritionnelles de matières premières végétales (céréales, tourteaux, sous-produits), y compris pour des paramètres issus d'essais de digestibilité *in vivo*. Les valeurs de référence de digestibilité *in vivo* utilisées pour le développement des modèles NIR sont obtenues par calcul à partir de déterminations analytiques sur les aliments et/ou sur les matières fécales. Afin d'optimiser ce process et de gagner en efficience, il est proposé d'utiliser la spectroscopie proche infrarouge en substitution de ces méthodes analytiques primaires, plus longues et plus coûteuses.

Deux cas d'utilisation sont étudiés :

- Le premier concerne la détermination par NIR de l'énergie brute dans les matières fécales de poulet, nécessaire au calcul de l'énergie métabolisable (AME) des aliments / matières premières au lieu de la bombe calorimétrique.
- Le second porte sur la détermination par NIR de la composition en acides aminés dans les matières fécales de coqs, utilisé pour le calcul de la digestibilité des acides aminés dans les matières premières, en substitution aux analyses HPLC.

Dans chaque cas la prédiction NIR intervient dans le calcul de la valeur de référence *in vivo* (digestibilité ou AME) qui sera utilisée pour le développement des modèles PNE.

Dans un premier temps, la faisabilité de la prédiction par modèles NIR des paramètres d'énergie brute et de composition en acides aminés sur matière fécale est évalué. Dans un second temps, l'impact de la substitution des méthodes analytiques conventionnelles par le modèle NIRS sur le calcul final du paramètre de digestibilité (énergie métabolisable AME et Digestibilité des acides aminés) est analysé, afin de valider la pertinence de cette méthode alternative.

40 ans d'innovations pour la Planète : robots, SPIR, IA…

Antoine BOURELY [1]

[1] *ABY Circular, La Tour d'Aigues, France*
Email : antoine.bourely@outlook.fr

***Mots-clefs** : robotique, spectroscopie, recyclage*

Cette présentation est une rétrospective de ma carrière, commencée dans la recherche publique au CEMAGREF, et poursuivie dans une première PME française devenue ETI, Pellenc SA, équipementier agricole, puis dans une jeune pousse de la même origine, Pellenc ST, fournisseur de tri optique, que j'ai co-fondée en 2001 et qui a atteint 300 personnes en 2025.

Ayant pris ma retraite début 2026, je garde un statut de consultant, pour continuer d'accompagner le développement de Pellenc ST.

Ma vision de départ était déjà claire, agir pour l'environnement par la technologie, mais je n'avais aucune idée de la forme que cela prendrait. Mon parcours a fait un long détour par l'agriculture, quand l'environnement n'avait pas de financements. Puis, je suis revenu vers l'écologie quand l'occasion s'est présentée avec le développement du recyclage. J'ai pu exploiter les synergies entre ces deux mondes, de mentalités pourtant très différentes. Et je me suis retrouvé à trier des poubelles partout dans le Monde…

Passionné de technologies, mais ouvert et pas trop spécialisé, je suis passé successivement par la robotique, puis la vision industrielle, l'optique et la spectroscopie infra-rouge, pour finir sur l'intelligence artificielle et les marqueurs invisibles : tout ce qui peut contribuer au résultat est digne d'intérêt.

J'ai rencontré des mentors et de nombreux collègues passionnés, et toujours gardé de bonnes relations avec eux, même après avoir changé de structure. J'ai réalisé par ma personne un transfert de la recherche vers les applications industrielles, mais toujours en équipe : on ne réussit rien tout seul.

La présentation montrera aussi que rien n'est gagné d'avance, que le parcours est parsemé d'échecs, qu'il faut savoir surmonter avec persévérance et toujours une vision du but final. Le succès est aussi le résultat d'interventions clés des décideurs dans les phases critiques, et notamment Francis Sévila et Roger Pellenc.

J'espère que ce parcours donnera des idées aux jeunes innovateurs, et à tous ceux qui se battent pour la Planète.

Vers un modèle de fondation pour la spectroscopie proche infrarouge… ou pas

Maxime Metz[1,3,4,5], Florent Abdelghafour[2,3,5], David Estève[1,3], Magalie Claeys-Bruno[4,5]

[1]Pellenc ST, Applied Research Group, 84120 Pertuis, France
[2]ITAP, Univ. Montpellier, INRAE, Institut Agro, 34196 Montpellier, France
[3]LabCom Aioly, Artificial Intelligence and Optics Laboratory, 34196 Montpellier, France
[4]Institut Méditerranéen de Biodiversité et d'Écologie marine et continentale, Aix-Marseille Université, UMR CNRS IRD Avignon Université, Marseille, France
[5]ChemHouse, Research Group, Montpellier, France

Email :m.metz@pellencst.com

***Mots-clefs** : spectroscopie proche infrarouge ; apprentissage profond ; modèles de fondation ; chimiométrie*

Les modèles de fondation ont récemment marqué une rupture importante dans plusieurs domaines de l'apprentissage automatique. En traitement du langage, en vision par ordinateur ou encore sur données tabulaires, des approches telles que GPT, DINOv2/DINOv3 ou TabPFN ont montré l'intérêt d'apprendre, à partir de grands volumes de données, des représentations générales pouvant ensuite être réutilisées pour différentes tâches avec une adaptation limitée [1–4]. Cette logique ouvre des perspectives intéressantes pour la spectroscopie proche infrarouge (NIR), où les jeux de données disponibles sont souvent de taille réduite, hétérogènes, acquis avec des instruments différents et associés à des propriétés chimiques ou physiques variées.

Dans ce travail, nous explorons cette question à travers l'évaluation d'un modèle profond entraîné conjointement sur plusieurs bases de données NIR, avec l'objectif d'étudier sa capacité à partager de l'information entre jeux de données plutôt que d'apprendre indépendamment chaque problème. L'étude est conduite sur 21 tâches de régression issues de bases de données de chimiométrie et de spectroscopie disponibles dans JchemoData.jl [5]. Ces tâches couvrent des situations volontairement diverses, avec des effectifs, des propriétés cibles, des résolutions et des gammes spectrales différentes, ainsi que des relations spectre-propriété pouvant être plus ou moins linéaires. Les résultats obtenus restent contrastés. Le modèle proposé ne surpasse pas globalement la PLS, qui demeure une référence particulièrement performante pour des problèmes peu complexes, notamment lorsque le nombre d'échantillons est limité. Il présente néanmoins de meilleures performances sur certaines bases, suggérant qu'un apprentissage multi-bases peut capter des régularités utiles dans des contextes spectroscopiques hétérogènes.

Ces premiers résultats ne constituent donc pas une démonstration de supériorité des approches profondes sur les méthodes chimiométriques classiques, mais plutôt un premier pas vers des modèles spectraux plus génériques. Ils soulignent également les défis restant pour développer un véritable modèle de fondation en spectroscopie NIR, en particulier la prise en compte explicite de la structure physique du spectre, des différences entre instruments, des prétraitements spectroscopiques et de la diversité des données disponibles.

References

[1] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever. *Improving Language Understanding by Generative Pre-Training*. OpenAI, 2018.
[2] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al. *DINOv2: Learning Robust Visual Features without Supervision*. arXiv:2304.07193, 2023.
[3] O. Siméoni, H. Vo, M. Seitzer, V. Baldassarre, M. Oquab, M. Berman, T. Darcet, T. Moutakanni, et al. *DINOv3*. arXiv:2508.10104, 2025.
[4] N. Hollmann, S. Müller, K. Eggensperger, F. Hutter. *TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second*. arXiv:2207.01848, 2022.
[5] M. Lesnoff. *JchemoData.jl: Chemometrics and Spectroscopy Datasets*. GitHub repository.

Chimiométrie profonde & détection d'anomalie : apprentissage autosupervisé pour l'authentification de produit alimentaire

Florent Abdelghafour [1,3,4], Daniele Tanzilli [1,3,4], Jean-Michel Roger[1,3,4], Maxime Metz[2,3,4,5]

[1] *ITAP, Univ. Montpellier, INRAE, Institut Agro, 34196 Montpellier, France*
[2] *Pellenc ST, Applied Research Group, 84120, Pertuis, France*
[3] *ChemHouse, Research Group, Montpellier, France*
[4] *LabCom Aioly, Artificial Intelligence and Optics Laboratory, 34196, Montpellier, France*
[5] *Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile, Marseille, France*

Email : florent.abdelghafour@inrae.fr

***Mots-clefs** : Apprentissage auto-supervisée, détection d'anomalie, apprentissage profond, qualité sécurité alimentaire*

Les problèmes d'authentification et de détection de fraude peuvent être formulés comme une question de conformité, consistant à déterminer si un produit appartient à l'espace des échantillons authentiques ou s'en écarte. Dans ce contexte, la modélisation en une classe vise à caractériser la variabilité des produits authentiques en utilisant uniquement des échantillons authentiques. Ce cadre est particulièrement pertinent en sécurité alimentaire et en contrôle qualité, où l'ensemble des adultérations possibles ou des dérives de procédé ne peut être représenté de manière exhaustive lors de la phase de calibration. Cependant, les méthodes standards comme SIMCA ainsi que des extensions récentes basées sur l'apprentissage profond, telles que VAE-SIMCA [1], font face à une difficulté fondamentale, car elles nécessitent un jeu de données volumineux et hautement représentatif d'une seule classe spécifique afin de définir une frontière robuste. En pratique, de tels ensembles de calibration sont difficiles à obtenir, ce qui conduit à des modèles qui soit surajustent la distribution de référence en rejetant à tort des échantillons légitimes, soit deviennent trop permissifs et échouent à détecter les anomalies.

Pour répondre à cette dépendance à des données massives spécifiques à une classe, un cadre d'apprentissage profond auto-supervisé est introduit. Cette approche exploite un transformeur spectral pré-entraîné à l'aide de tâches de masquage et de corruption afin de reconstruire les régions manquantes des spectres. L'intérêt principal de cette stratégie réside dans sa capacité à apprendre les structures spectrales intrinsèques à partir de jeux de données divers avant d'être appliquée à une tâche spécifique de conformité. En extrayant des caractéristiques de haut niveau dans un espace métrique qualitatif, le modèle atteint une meilleure représentativité même lorsque les échantillons de la classe cible sont limités. Cet espace latent structuré permet ensuite la mise en œuvre de stratégies robustes de détection d'anomalies, telles que les k plus proches voisins (k-NN), où une observation est évaluée en fonction de sa densité locale et de sa proximité avec des motifs authentiques connus, plutôt que sur de simples métriques de reconstruction. Cette stratégie est comparée aux approches conventionnelles basées sur SIMCA et sur des VAE de type CNN appliquées à des données de spectroscopie infrarouge, montrant comment le pré-entraînement auto-supervisé réduit la sensibilité aux limitations du jeu de calibration en fournissant une représentation plus complète de la variabilité spectrale.

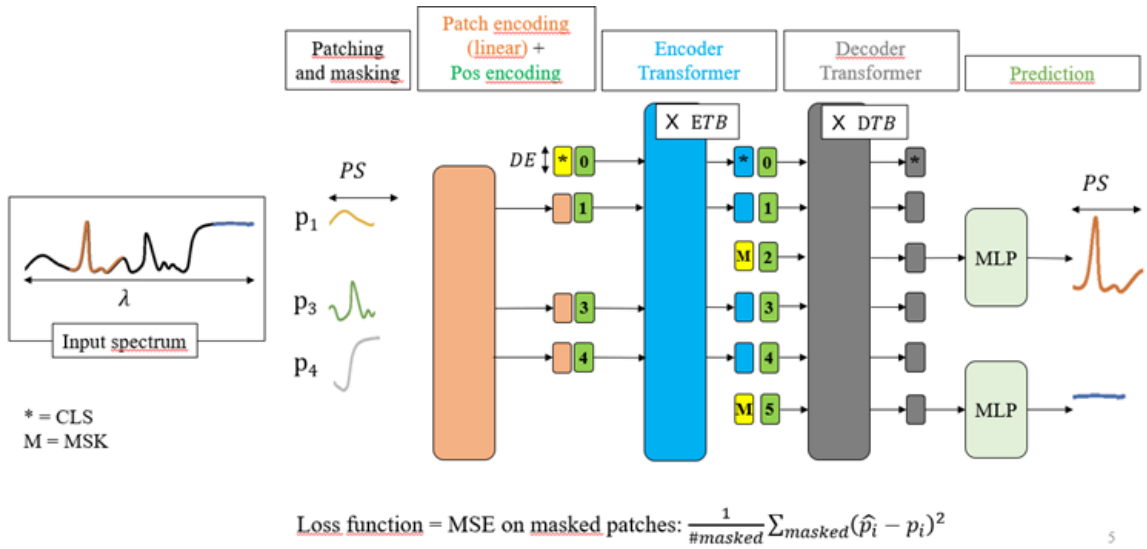


Figure 1. Tâche de pré-apprentissage auto-supervisé (SSL) : reconstruction des parties manquantes d'un spectre afin d'obtenir un espace latent plus générique.

Remerciements

Ce travail a été soutenu par la subvention n° ANR-23-LCV2-0015, dénommée ANR AIOLY. Ce travail a bénéficié d'un accès aux ressources de calcul haute performance (HPC) de l'IDRIS dans le cadre de l'allocation 2023-[AD010114820] accordée par GENCI.

References

[1] Akam Petersen, Sergey Kucheryavskiy, VAE-SIMCA — Data-driven method for building one class classifiers with variational autoencoders, Chemolab.,2025,https://doi.org/10.1016/j.chemolab.2024.105276.

[2] M. Wan et al., « MAE-NIR: A masked autoencoder that enhances near-infrared spectral data to predict soil properties », Compag, doi: 10.1016/j.compag.2023.108427.

Attention-based aggregation of preprocessing for spectral deep learning

Daniele Tanzilli[1,2], Florent Abdelghafour[1,2], Maxime Metz[1,3,4]

[1] *LabCom Aioly, Artificial Intelligence and Optics Laboratory, 34196, Montpellier, France,*
[2] *ITAP, Univ. Montpellier, INRAE, Institut Agro, 34196, Montpellier, France,*
[3] *Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile Marseille, France,*
[4] *Pellenc ST, Applied Research Group, 84120, Pertuis, France*

Email : daniele.tanzilli@inrae.fr

***Keywords**: Deep chemometrics; spectral pre-processing; attention mechanism; convolutional neural networks; design of experiments.*

Introduction

Spectroscopic data are widely used in analytical chemistry, but their effective modelling remains challenging due to noise, instrumental artefacts and measurement-related variability. In chemometrics, spectral pre-processing is therefore a crucial step to enhance chemically relevant information and reduce unwanted sources of variation [1]. At the same time, deep learning models, particularly convolutional neural networks, have shown strong potential for modelling complex and nonlinear spectral data. However, the systematic integration of chemometric pre-processing knowledge into deep learning architectures remains only partially explored. In this work, we propose PASTIS-Net, a compact deep learning architecture designed to exploit multiple chemometric pre-processing strategies without excessively increasing model complexity. Instead of concatenating differently pre-processed spectra along the wavelength dimension [2], PASTIS-Net organises them as separate input channels. This preserves the original spectral length and allows the network to learn from complementary spectral representations more efficiently. A convolutional embedding step is followed by a self-aggregation mechanism based on attention, which adaptively weights the contribution of each pre-processed representation and promotes the selection of the most informative spectral features.

Materials and methods

The proposed methodology was evaluated using the Open Soil Spectral Library (OSSL) [3], which is an open-access dataset comprising over 155,500 soil spectra samples from a variety of soil types and geographic regions. The database includes visible-near-infrared and mid-infrared spectra acquired using different spectrometers. It also provides laboratory-measured soil properties, such as organic carbon content. PASTIS-Net was trained on this dataset and compared with existing deep learning approaches.

Resultas and discussion

The results show that providing the network with multiple pre-processed spectral representations allows PASTIS-Net to achieve improved predictive performance while keeping the architecture compact. By organising the different pre-processing outputs as input channels, rather than concatenating them along the spectral dimension, the model can exploit complementary chemical information without increasing the input size handled by the subsequent layers. As a result, PASTIS-Net requires a substantially lower number of trainable parameters compared with reference deep learning architectures.

The attention-based aggregation mechanism further improves this strategy by adaptively weighting the contribution of each pre-processed representation. This enables the network to focus on the most informative transformations and to limit the impact of less relevant or non-optimal pre-processing

methods. Overall, the use of multiple pre-processing inputs makes it possible to design more efficient spectral deep learning models, combining improved predictive ability with reduced model complexity.

References

[1] Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy–when, why and how. Analytical Methods, 6(18), 7124-7129.

[2] Mishra, P., & Passos, D. (2021). A synergistic use of chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit. Chemometrics and Intelligent Laboratory Systems, 212, 104287.

[3] Zhou, L., Zhang, C., Taha, M. F., Wei, X., He, Y., Qiu, Z., & Liu, Y. (2020). Wheat kernel variety identification based on a large near-infrared spectral dataset and a novel deep learning-based feature selection method. Frontiers in plant science, 11, 575810.

Didactical use of PLS, Local PLS, SVM and ANN regressions: stevia and soybean

Jean Brustel [1,2*], Hélène Missonnier [1,2], Cyril Libourel [1,2], Maria Angelica Usta[3], François Perdrieux [1,2], Amandine Chrétien [1,2], Lauriane Heissat [1,2], Sarah Happiette [3], Monique Berger [1,2], Cécile Levasseur-Garcia [3].

[1] Université de Toulouse, Ecole d'Ingénieurs de Purpan, Laboratoire d'Analyses Physico-chimiques (LAP), Toulouse, France
[2] Université de Toulouse, École d'ingénieurs de PURPAN, UPR - Physiologie, Pathologie et Génétique Végétales (PPGV), Toulouse, France
[3] Université de Toulouse, École d'ingénieurs de PURPAN, Unité de recherche OCCI'FOOD, Toulouse, France

**Email : jean.brustel@gmail.com*

***Keywords**: Soybean, Stevia, Specialized metabolites, PLS, Nonlinear regression.*

Abstract:

Near-infrared spectroscopy (NIRS) is an economical and non-destructive tool for rapidly characterizing the physicochemical properties of agricultural crop products. The amount and composition of specialized metabolites are essential for determining the quality of these products, yet their low concentration requires advanced methods for accurate prediction. The content and composition of steviol glycosides in stevia leaves and isoflavones in soybeans were investigated. The stevia plant material was grown and analyzed at the Ecole d'Ingénieurs de Purpan and includes 300 samples. For soybeans, the material and data were produced and processed during the SOYFOOD+ project and the ANR SOYSTAINABLE and counts 750 samples. The spectra were collected using a Bruker MPA II (reflectance mode, 800 – 2800nm, 0.25 nm step, 64 scans per analysis), metabolic compounds were measured by HPLC method. Spectral outliers were detected using Hotelling's T^2 and Q residuals method (limit acceptability set at 5%) and removed from the databases. Kenard-Stone sampling was used to distinguish the calibration set (80%) and the validation set (20%) for the models. For soybeans, an additional set of 100 samples grown under distinct conditions (year and location) will serve as external validation. The spectra are smoothed using moving averages, and several preprocessing methods were evaluated: raw spectra, first and second Stavisky-Golay derivatives, SNV transformation associated with detrending and MSC. The performances of PLS linear models were evaluated, as well as nonlinear approaches such as local-PLS, SVM, and ANN. For each model, combinations of preprocessing methods with hyperparameters were evaluated. The risk of model overfitting was also monitored.

DROP-DA: A discriminant model that avoids overfitting

Jean-Michel Roger [1,2]

[1] *IINRAE UMR ITAP, Montpellier, France*
[2] *ChemHouse, Research Group, Montpellier, France*

Email : jean-michel.roger@inrae.fr

***keywords** : Multivariate discrimination, NIRS*

Introduction

Linear discriminant analysis remains a valuable tool for supervised classification, but its application to highly multivariate data such as spectroscopy is often challenging. Several approaches have been proposed to overcome these limitations, including the use of PCA or PLS as preprocessing steps [2], or continuum methods that establish a compromise between PLS and LDA [3].

This work presents DROP-DA [4], a discriminant analysis method based on orthogonal subspace projection. The approach first removes a part of the variability associated with within-class variation and then constructs discriminant directions from the remaining information. The degree of filtering is controlled by a parameter that can be optimized by cross-validation.

Material and methods

The method was evaluated on several spectroscopic datasets and compared with PLS-DA. The comparison considered both classification performance and the interpretability of the resulting discriminant vectors.

Results and discussion

The results show that:

- DROP-DA does not always reach the predictive performance of PLS-DA
- DROP-DA is unsensitive to overfitting
- DROP-DA provides discriminant vectors that are generally easier to interpret, making it a valuable alternative when understanding the variables responsible for class separation is an important objective

References

[1] Fisher, R.A. (1936). *The use of multiple measurements in taxonomic problems*. Annals of Eugenics, 7, 179–188.

[2] Barker, M., & Rayens, W. (2003). *Partial least squares for discrimination*. Journal of Chemometrics, 17, 166–173.

[3] Nocairi, H., Hanafi, M., & Qannari, E.M. (2005). *Approche continuum de la discrimination de type ridge*. Revue de Statistique Appliquée, 53(2), 29–41.

[4] Hadoux, X., Rutledge, D. N., Rabatel, G., & Roger, J. M. (2015). DROP-D: dimension reduction by orthogonal projection for discrimination. *Chemometrics and Intelligent Laboratory Systems*, *146*, 221-231.

Diagnostic nutritionnel en micro et macronutriments des limbes, pétioles et baies de raisin de *vitis vinifera* par l'utilisation du proche-infrarouge et d'outils de chimiométrie

Sébastien Cuq[1], Valérie Lemetter[1], Cécile Levasseur-Garcia[1]

[1] *University of Toulouse, PURPAN Engineering School, Occi'Food Research Unit, Toulouse, France*

Email : sebastien.cuq@gmail.com

***Mots-clefs** : Vitis vinifera, nutrition, macronutriments, micronutriments, spectroscopie, chimiométrie*

Les macronutriments (phosphore [P], potassium [K], calcium [Ca] et magnésium [Mg]), ainsi que les micronutriments (manganèse [Mn], fer [Fe], cuivre [Cu], zinc [Zn] et bore [B]) jouent un rôle essentiel, pas uniquement dans la croissance physiologique de la vigne, mais aussi sur la qualité raisin, et donc du vin, qui est produit. Le taux de chaque nutriment est généralement déterminé à partir du limbe ou du pétiole des feuilles. Entre les prélèvements et la réception des résultats, peut s'observer un délai habituel de deux semaines, ce qui limite l'intervention en temps réel des viticulteurs pour ajuster la fertilisation. La spectroscopie proche-infrarouge et les outils chimiométriques sont une réponse possible à la réduction des délais d'analyse.

Dans cette étude, 67 parcelles de vignes ont été suivies sur plusieurs campagne de culture, avec prélèvements de limbes, pétioles, baies petit-pois et baies à véraison, à différents stades physiologiques des plantes. Les spectres ont été acquis sur matériels frais et secs broyés, et des analyses ont permis de rapprocher ces spectres aux taux de P, K, Ca, Mg, Mn, Fe, Cu, Zn et B contenu dans les échantillons. Les 677 spectres ont ensuite été utilisé pour développer des modèles de régression PLS, pour obtenir une approche quantitative (dans le meilleur des cas) ou permettre une approche de classification (carence / nutrition normal / excès). La performance des modèles a été évaluée en se basant sur le RPD (ratio of performance to deviation).

Les résultats ont été bien meilleurs sur les échantillons secs broyés, montrant des capacités prédictives pour le Ca (RPD > 2), et des classifications possibles pour le K, Mg, Mn, Fe, Zn et Cu (RPD compris entre 1,4 et 2). Pour le P et le B, les modèles ne se sont pas montrés assez précis.

Transfert d'étalonnages NIR du diagnostic foliaire du palmier à huile pour le déploiement d'un réseau de micro-spectromètres au champ

Sophie Kamgaing[1,2,3], Sylvain Vrignon[4,5], Antoine Versini[6] , Gilles Chaix[1,2,3]

[1] *CIRAD - UMR AGAP Institut, Montpellier, France*
[2] *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*
[3] *ChemHouse Research Group, Montpellier, France*
[4] *CIRAD, UMR ABSys, Montpellier, France*
[5] *UMR ABSys, Univ. Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, Montpellier, France*
[6] *CIRAD UPR Recyclage et risque, Montpellier, France*

Email : sophie-kamgaing .simo@cirad.fr, sophiekasisopa@gmail.com

***Mots-clefs** : NIRS, transfert de calibration, chimiométrie*

La fertilisation du palmier à huile repose sur un apport optimal d'éléments minéraux du sol, à la croissance de la plante et à l'obtention d'une production accrue et durable. Les éléments nutritifs tels que l'azote, le phosphore, le potassium, le manganèse, le calcium ou encore le magnésium jouent un rôle essentiel dans le développement du palmier. Pour ajuster ces apports, l'analyse foliaire réalisée en laboratoire constitue aujourd'hui la méthode de référence. Cependant, son coût élevé et son accessibilité limitée dans de nombreuses régions tropicales freinent son utilisation par les producteurs, en particulier par les petits planteurs. Dans ce contexte, les équipes de recherche du CIRAD ont démontré l'intérêt de la spectroscopie proche infrarouge (NIR), aussi bien en laboratoire que directement sur le terrain, pour quantifier les éléments minéraux des feuilles [1]. Des modèles de prédiction ont ainsi été développés à partir de spectres NIR acquis sur différents appareils sur feuilles sèches, notamment le Tango (Bruker,800–2500 nm) et le NirOne (Spectral Engines, 1750– 2150 nm), associés aux mesures de référence des teneurs en N, P, K, Ca, Mg.

L'objectif de cette étude est de réaliser un transfert d'étalonnage entre plusieurs spectromètres, en particulier du Tango vers NirOne, afin de pouvoir réaliser des prédictions des teneurs minérales sur sites (bord champs).

Pour construire les modèles prédictifs, nous avons appliqué la régression PLS (Partial Least Squares), qui permet d'estimer les valeurs de référence à partir des spectres mesurés. Toutefois, la robustesse du modèle peut être compromise lorsque les spectres proviennent d'appareils différents, d'où la nécessité d'un transfert de calibration. Pour cela, plusieurs méthodes de correction spectrale ont été testées : Update, double Update, PDS (Piecewise Direct Standardization) et EPO (External Parameter Orthogonalisation) qui sont des méthodes de transfert d'étalonnage utilisées classiquement en proche infrarouge d'un instrument source vers un spectromètre cible [2].

La base d'étalonnage (spectres Tango = source, sur la gamme et à la résolution spectrale du NirOne) a été divisée en jeu de calibration (435 spectres) et jeu de validation (31 spectres) à l'aide de la méthode Duplex, garantissant une répartition représentative. Les spectres de ces mêmes échantillons de validation mesurés sur le spectromètre NirOne ont permis d'évaluer l'efficacité des différentes méthodes de transfert. Des échantillons standards (62) ont été mesurés sur les deux appareils et ont été divisés en 2 jeux de standards, le premier pour créer la matrice des perturbations, et le second pour régler les paramètres de certaines méthodes.

Dans un premier temps, les modèles PLS ont été calibrés avec le package rchemo [3] sur les données du Tango, une validation croisée k-fold (k = 2, 20 répétitions) a permis de sélectionner le nombre optimal

de variables latentes en minimisant le RMSECV et le meilleur prétraitement. Les performances ont ensuite été évaluées sur les spectres, du Tango et du NirOne, des échantillons de validation, via le RMSEP. Les résultats ici présentés ne concerne que la teneur en azote.

Pour les méthodes UPDATE et double Update, nous avons rajouté à la base d'étalonnage du Tango, les spectres du jeu 1 des standards du NirOne, et du Tango pour la seconde respectivement. Pour la PDS et l'EPO nous avons corrigé les spectres de la base d'étalonnage. Nous avons comparé les résultats en prédiction des spectres du jeu de validation du NirOne pour comparer l'efficacité des différentes méthodes sur le RMSEP.

Avant la mise en œuvre des méthodes de transfert, l'étalonnage a été réalisé à partir des spectres acquis avec l'instrument Tango. La validation croisée de ce modèle a conduit à une erreur RMSECV de 0,074 % DM, tandis que la prédiction du jeu de test Tango a donné une erreur RMSEP de 0,116 % DM. En revanche, lorsque ce même modèle a été appliqué au jeu de test acquis avec l'instrument NirOne, l'erreur de prédiction (RMSEP) a atteint 0,398 % DM mettant en évidence l'impact des différences instrumentales et la nécessité de recourir à des méthodes de transfert de calibration.

Les méthodes Update et Double Update ont conduit à des valeurs de RMSEP relativement proches, avec des résultats respectifs de 0,297 % DM et 0,276 % DM. En conséquence, les approches Update et Double Update n'ont pas permis d'améliorer significativement la qualité des prédictions.

La méthode PDS, appliquée au jeu de test NirOne après prétraitement, a montré une erreur en test de 0,196 % DM. De son côté, la méthode EPO appliquée dans les mêmes conditions, a conduit à une erreur de 0,233% DM.

À l'issue de cette étude, la méthode PDS apparaît actuellement comme la plus performante parmi celles qui ont été testées. Les résultats en test après correction PDS peuvent paraître encore élevés, mais il faut les relativiser du fait que les deux spectromètres sont très différents au niveau de la technologie, de l'optique, etc. Toutefois, d'autres méthodes de machine Learning seront explorées afin d'évaluer leur capacité à améliorer ou non les transferts d'étalonnage.

Références

[1] Avit V. (2023). Impact de l'appareil de mesure utilisé et de la préparation des échantillons sur des modèles de diagnostic foliaire du palmier à huile par spectrométrie proche infrarouge. Rencontres HelioSPIR 2023.
[2] Wang, Yongdong, Veltkamp, David J., Kowalski, Bruce R. (1991) Instrument standardization.
Analytical Chemistry, 63 (23). 2750-2756 doi:10.1021/ac00023a016
[3] Brandolini-Bunlon M., Jallais B., Roger J.M. Lesnoff M.,2023 R package rchemo: Dimension reduction, regression and discrimination for chemometrics. https://github.com/Chemhouse/group/rchemo

Comparaison des approches de régression PLS et d'apprentissage automatique pour la caractérisation chimique de l'huile de palme rouge par spectroscopie proche infrarouge

Matéo Rosa[1,2,3,*], Stéphane Dussert[2], Virginie Vaissayre[2], Amel Joualli[2], Antoine Giordani[2], Thibaut Bontpart[2,3], Hubert Domonhédo[4], Florence Jacob[5], Fabienne Morcillo[2,3] , Martin Ecarnot[1]

[1] *INRAE, ARCAD, Montpellier, France*
[2] *DIADE, Univ Montpellier, CIRAD, IRD, Montpellier, France*
[3] *CIRAD, DIADE, F-34398 Montpellier, France*
[4] *CRA-PP/INRAB, Benin*
[5] *PalmElit SAS, Montferrier-sur-Lez, France*
[*] *Institut Agro Rennes-Angers (anciennement AgroCampusOuest), France*

Email : mateo.rosa@agrocampus-ouest.fr

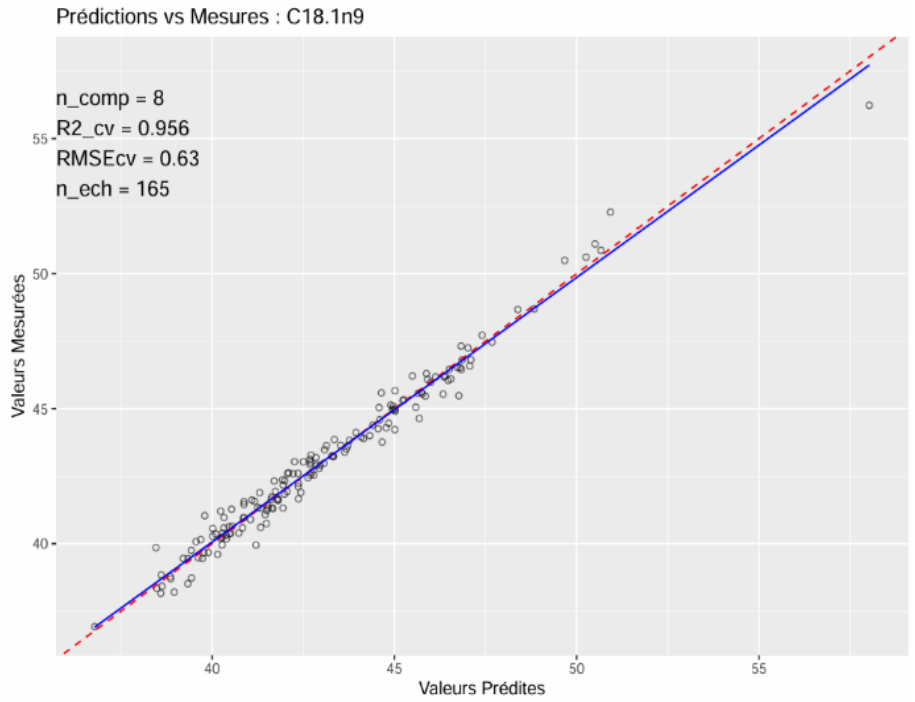
***Mots-clefs** : Huile rouge, NIRS, PLSr, machine learning, prétraitement, RandomForest*

Il existe une grande diversité entre les différentes origines africaines du palmier à huile pour la teneur et la composition en acides gras et vitamines de l'huile rouge extraite des fruits. Afin d'identifier les déterminants génétiques et environnementaux qui gouvernent cette diversité, une méthode de phénotypage rapide par spectroscopie proche infrarouge (SPIR) est en cours de développement pour les différentes familles de composés (projet VitaSPEC).

Au total, environ 600 échantillons de mésocarpe de fruit et 470 échantillons d'huile ont été collectés, pour plus de 36.000 spectres capturés. Pour les fruits, les spectres ont été acquis selon 3 modalités différentes : mésocarpe frais, mésocarpe séché sur silicagel, et mésocarpe lyophilisé. En parallèle, des analyses chimiques ont été réalisées sur 168 échantillons d'huile rouge et 160 échantillons de fruits, pour quantifier les acides gras, les carotènes (provitamines A) et les tocochromanols (vitamines E).

Les données spectrales et chimiques ont dans un premier temps été traitées indépendamment pour chaque composé cible par l'approche classique de régression des moindres carrés partiels (PLS). Nous avons testé une très large gamme de prétraitements et, pour chaque composé, affiné le meilleur ajustement des paramètres. De très bons résultats sont obtenus en validation croisée pour certains composés (ex : acide oléique 18:1n-9 ; R2CV > 0,95 ; Fig. 1). D'autres composés sont estimés avec une moins bonne précision (ex : teneur en carotènes ; R2CV ≈ 0,75).

Figure 1 : Régression PLSr et performance pour l'acide oléique (C18:1n-9)



Dans un second temps, plusieurs algorithmes de Machine Learning non-linéaire (notamment Forêts Aléatoires) ont été testés. Leur performance sera comparée dans cette présentation à celle des approches de régression PLS.

Near-Infrared Spectroscopy for wood species discrimination and cardanol binder detection in biomass pellets

Luiza Mendonça Bonfim Tavares [1*], Douglas Lamounier Faria [1], Thiago de Paula Protásio[1], Gilles Chaix[2,3,4], Paulo Ricardo Gherardi Hein[1]

[1] *Department of Forest Science - UFLA, Lavras, MG - Brazil*
[2] *CIRAD - UMR AGAP Institut, Montpellier, France*
[3] *UMR AGAP Institut, University of Montpellier, CIRAD, INRAE, Agro Institute, Montpellier, France*
[4] *ChemHouse, Research Group, Montpellier, France*

Email : luiza.tavares3@estudante.ufla.br, douglas.faria3@ufla.br, thiagoprotasio@ufla.br, gilles.chaix@cirad.fr, paulo.hein@ufla.br

***Mots-clefs** : Multivariate classification, forest biomass, bio-binder*

Introduction

The main cause of current climate change can be attributed to greenhouse gas (GHG) emissions. A large part of CO2 emissions released into the atmosphere come from the burning of fossil fuels, mainly coal, oil and gas [1]. A sustainable alternative for generating clean energy is the use of biomass. Forest residues such as branches, leaves, and small trees have great potential for this purpose [2], however, the use of lignocellulosic biomass as a renewable energy source is limited by its high moisture content and low energy density. These problems can be minimized with palletization [3]. The addition of binders, such as organic and inorganic ones, improves both the physical characteristics, such as hardness and mechanical resistance, and the combustion properties and calorific value of the pellets [4], but their use is regulated by standards (ENplus and ISO 17225-2) [5]. Due to potential unwanted gas emissions and the need to control ash content, near-infrared (NIR) spectroscopy emerges as a fast, non-destructive, and in-line applicable alternative. Its real-time measurement capability and process optimization throughout the production chain allow its use in monitoring material quality [6]. This study evaluated the potential of NIR to discriminate wood species in pellet production and to detect cardanol in the same samples.

Materials and methods

Waste from eight agroforestry timber species (Cordia sp., Bowdichia virgilioides Kunth, Parkia pendula (Willd.) Benth. ex Walp., Homalolepis cedron (Planch.) Devecchi & Pirani, Ocotea sp., Apeiba echinata Gaertn., Tapirira guianensis Aubi., Protium sp.), were compacted with and without the binder cardanol using an Eng-Maq® 0200v pelletizer, with a compaction ratio of 3.33. Pelletizing was carried out at a temperature between 80 and 95 °C and a pressure of 29.42 MPa. The natural binder cardanol was used at a concentration of 2% (dry mass basis of the material), the limit stipulated by the ISO 17225-2 (2020) standard [5]. Pellets without the addition of cardanol were produced as a control for comparison regarding the final quality of the pellets. These were analyzed on a Fourier transform (FT) NIR spectrometer (MPA, Bruker Optik GmbH, Ettlingen, Germany) with its OPUS v. 7.0 software. The spectral range used for analysis was 1100 – 2500 nm (9000 – 4000 cm^{-1}) with a resolution of 8 cm^{-1}, resulting in 1300 spectral variables. Sixteen scans per spectra were performed and averaged. Background compensation was performed for each species. Spectral data were analyzed using Principal Component Analysis (PCA) for exploratory analysis and Partial Least Squares Regression (PLS-DA) for classification. Data analysis was performed with the rchemo [7] package and Rstudio. NIR spectra were evaluated in their raw (unprocessed) form and after applying first-order Savitzky-Golay derivatives (1D) (15 smoothing points, second-order polynomial).

Results and discussion

The NIR spectral profiles of the pellets were typical of lignocellulosic materials, with variations in shape and intensity reflecting the species composition and the addition of cardanol. Data preprocessing accentuated these distinctions, mitigating scattering and baseline interference. PCA analysis showed that the first principal components explained most of the variance, revealing clear groupings by species and, albeit subtler, separation trends regarding the presence of cardanol. The PLS-DA models achieved high performance, being able to identify the cardanol signature without compromising the botanical classification of the samples. Therefore, the combination of NIR spectroscopy with PLS-DA constitutes an effective, agile, and non-invasive technique for identifying species and the cardanol additive in biomass pellets.

Acknowledgments

The authors gratefully acknowledge the financial support from the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Finance Code 001; the National Council for Scientific and Technological Development (CNPq, grant no. 409701/2024-6); the Minas Gerais State Research Support Foundation (FAPEMIG, grant nos. APQ-00742-23, APQ-03095-26, APQ-05808-26, APQ-05431-24, and APD-00296-25); and TO 032/2026 – INOV@ÇÃO UFLA (Process no. 23090.000734/2026-70). P.R.G. Hein was supported by CNPq grant no. 304353/2024-8. The authors also thank the French Agricultural Research Centre for International Development (CIRAD) for the institutional support (D2S grant) and technical assistance.

References

[1] SANG, Huiying et al. High-value utilization of biomass pellets: Paradigm shift from alternative fuel to multifunctional chemical platform—A review. Carbon Capture Science & Technology, p. 100601, 2026. https://doi.org/10.1016/j.ccst.2026.100601

[2] NAMAKKA, Murtala; RAHMAN, Md Rezaur; BAKRI, Muhammad Khusairy Bin. Waste biomass pellets for green energy production-A sustainable alternative for energy security. Renewable and Sustainable Energy Reviews, v. 226, p. 116315, 2026. https://doi.org/10.1016/j.rser.2025.116315

[3] PITAK, Lakkana et al. Predicting the true density of commercial biomass pellets using near-infrared hyperspectral imaging. Artificial Intelligence in Agriculture, v. 6, p. 266-275, 2022, https://doi.org/10.1016/j.aiia.2022.11.004

[4] SYKOROVA, Veronika et al. Binders in biomass pelletization and combustion analysis. Energy Conversion and Management, v. 357, p. 121473, 2026. https://doi.org/10.1016/j.enconman.2026.121473

[5] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 17225-2:2021: Solid biofuels — Fuel specifications and classes — Part 2: Graded wood pellets. Geneva: ISO, 2021

[6] POSOM, Jetsada; MARAPHUM, Kanvisit. Fast prediction of the combustion properties of biomass pellets using hyperspectral imaging. Biomass and Bioenergy, v. 183, p. 107134, 2024. https://doi.org/10.1016/j.biombioe.2024.107134

.

Impact of water stress on wood formation for *Eucalyptus grandis* as revealed by NIR spectroscopy

Nathália Cardoso Pereira[1], Gilles Chaix [2,3,4], Mario Tomazello-Filho[1]

[1] *ESALQ/USP, College of Agriculture "Luiz de Queiroz", Piracicaba, Brazil*
[2] *CIRAD - UMR AGAP Institut, Montpellier, France*
[3] *UMR AGAP Institut, University of Montpellier, CIRAD, INRAE, Agro Institute, Montpellier, France*
[4] *ChemHouse, Research Group, Montpellier, France*

Email: nathaliacp@usp.br

***Keywords**: Rainfall exclusion, Wood quality, Wood properties.*

Introduction

The gender *Eucalyptus* occurs naturally in Australia and was introduced to Brazil to achieve the demand for timber for railroad constructions; also, its adaptation occurred through an intensive process of genetic improvement, ex situ conservation, breeding and clonal selection. In 2025, the timber forestry sector hit a value of R$102.3 billion, with more than 8.1 million hectares planted with *Eucalyptus* spp., primarily for the pulp and paper industries [1]. At present, the greatest challenge facing the Brazilian forestry sector is understanding the species capacity to adapt to climate change; however, there are few studies addressing these changes in the properties of wood from trees in Brazil.

The use of hyperspectral data as a tool for detecting drought stress in plants is largely recognized and can provide information that enables the monitoring of the integrity of various biological materials, such as wood. Near-infrared spectroscopy (NIRS) detects and classifies various characteristics present in wood, such as its chemical components and physical properties, which can be altered under drought conditions during the growth and formation of wood.

In this project, a field experiment was conducted to simulate severe rain exclusion in *Eucalyptus grandis* trees in a planted forest area in Brazil, to identifying changes in wood quality. Within this context, the goal of this work was to select the best prediction model based on NIR spectral data, to identify water stress impacts in the wood of *Eucalyptus grandis*.

Materials and methods

The field experiment was installed in a plantation of half-sib progenies of *Eucalyptus grandis* trees at the Itatinga Forestry Science Experimental Station, São Paulo, Brazil. The experimental design was composed of two treatments, designated "Control" (C), with normal water conditions and "Rain Exclusion Treatment" (T) around 80%, which were monitored using automatic dendrometers fitted to the trees in three stages:

i. *Stage 0:* period from November 2014 to August 2015, represented by normal rainfall conditions;
ii. *Stage 1*: period from September 2015 to October 2018 (37 months), represented by an 80% rainfall reduction in T group, through the installation of transparent plastic sheeting fixed by wooden frames, where rainwater was collected and drained away from the plot. The sheeting was removed in October 2018;
iii. *Stage 2*: period from November 2018 to May 2023 (48 months), represented with the return to normal rainfall conditions.

One sample obtained from diameter at breast height (DBH) was taken by each tree and each stage was identified in wood samples, subsequently, near-infrared spectral (NIRS) data from 26 trees were collected using the ASD LabSpec 4 instrument. A multivariate statistical was used to analyze the spectral data; the raw spectra were corrected, and the first derivative method was used to smooth the spectral curves. Principal Component Analysis (PCA) was applied to explore the clustering of the spectra, and Partial Least Squares Discrimination Analysis (PLS-DA) was used for model selection to predict the types of wood according to stage. The cross-validation prediction errors were compared to selecting the best preprocessing spectra and optimal variable discriminant number, and the best model was selected based on the lowest error, these results were obtained using the '*rchemo*' package [3] with RStudio.

Results and Conclusion

The wood discs from two study groups evaluated by NIRS exhibited average spectral profiles similar to the properties that characterize the wood material. The similarity of the spectral signatures obtained from the PCA scores using the first derivative can be attributed to the chemical characteristics that characterize the wood, as well as to the different ages [3], throughout the study stages. The estimation models for the stages evaluated using PLS-DA based on the raw spectra were fitted, and the Savitzky-Golay treatment showed the lowest validation error. In the confusion matrix, the model achieved 100% accuracy in detecting stage 2 (normal conditions and mature trees), 96.6% accuracy in stage 1 (exclusion of rainfall) and 91.6% in stage 0 (normal conditions and early growth), with an overall model accuracy of 95.8% and a Kappa coefficient of 0.93. These results demonstrate the potential of NIRS analysis to distinguish and interpret the presence of water stress in the wood of *Eucalyptus grandis* trees.

Acknowledgements

The authors thanked the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Laboratory of Wood Anatomy and Identification (LAIM, ESALQ/USP), as well as CIRAD in Montpellier for the inter-institutional exchange facilitated by the D2S Strategic Support Programme – Hosting of PhD students.

References

[1] INDÚSTRIA BRASILEIRA DE ÁRVORES (Ibá). Relatório anual Ibá 2025. [S. l.]: Ibá, 2025. 100 p.

[2] Brandolini-Bunlon M., Jallais B., Roger J.M. Lesnoff M., 2023 R package rchemo: Dimension Reduction, Regression and Discrimination for Chemometrics, https://github.com/ChemHouse-group/rchemo.

[3] D. T. Medeiros, et al. Estimation of the basic density of *Eucalyptus grandis* wood chips at different moisture levels using benchtop and handheld NIR instruments, *Industrial Crops and Products*, v. 209, 117921, 2024, https://doi.org/10.1016/j.indcrop.2023.117921.

Distinction de maladies de conservation par NIRS sur pommes

Amandine Arnal[1,2], Céline Verdier[1], Sylvain Gerbaud[1], Léa Volmerange[2], Marielle Pages[2], Cecile Levasseur-Garcia[2]

[1] Absoger, 521 Chemin de la Graviere, 82100 Les Barthes, France
[2] Université de Toulouse, Ecole d'ingénieurs de PURPAN, Unité de recherche propre Occi'Food, Toulouse, France
Email : amandine.arnal@absoger.fr

***Mots-clefs** : NIRS, Pomme, Contamination, Maladies fongiques, Conservation*

L'industrie de la pomme doit répondre à des enjeux majeurs notamment dus aux maladies post-récolte, qui représentent une contrainte pour la qualité et la conservation des pommes. Ces pertes peuvent atteindre 3 à 12% selon les conditions de stockage et les pratiques associées. Ces dégâts sont causés pour 60 à 80% des cas par des pathogènes tels que *Penicillium spp., Botrytis cinerea, Phythophtora spp.* ou *Alternaria spp*. Ils ont une dynamique de développement qui dépend à la fois de facteurs biologiques et des conditions de stockage [1,2]. La détection de ces infections constitue un enjeu lors de la conservation en chambre froide qui dure plusieurs et dont l'accès est limité, en particulier dans un contexte industriel où les approches actuelles demeurent contraintes notamment par leur caractère destructif ou leur faible sensibilité [3].

La spectroscopie proche infrarouge (NIRS) a montré une bonne capacité pour la détection non destructive des défauts internes et des altérations physiologiques des fruits. Cette détection est possible par la sensibilité aux modifications de la composition biochimique et structurelle des tissus [3,4]. Une grande partie des travaux dans la littérature restent limitées à des conditions expérimentales contrôlées, souvent focalisées sur la distinction entre fruits sains et infectés. Il y a peu de travaux qui explorent la capacité du NIRS à discriminer différents types de maladies sur pomme dans des conditions proches du terrain[5,6].

Face à ce constat, l'objectif de cette étude est d'évaluer la capacité du NIRS à discriminer plusieurs maladies de conservation de la pomme. Des pommes de différentes variétés, issues du calibrage industriel et préalablement stockées à long terme sous atmosphère contrôlée, sont analysées. Ces fruits présentent diverses maladies de stockage. Les spectres sont acquis sur des zones asymptomatiques et symptomatiques d'un même fruit. Cela permet d'évaluer la discrimination de ces deux zones, tout en considérant l'effet de la variété et des conditions stockage.

Les résultats montrent une différence spectrale significative entre zones asymptomatiques et symptomatiques, observée sur l'ensemble des variétés, bien que l'intensité de cet effet puisse varier selon la variété. Une différenciation statistiquement significative entre les types de maladies est observée sur les zones symptomatiques, suggérant l'existence de signatures spectrales associées aux différents pathogènes. Cette structuration se traduit par une bonne capacité prédictive des modèles sur les zones symptomatiques (AUC > 0,9). Ces signatures pourraient être exploitées pour la discrimination des maladies. Une structuration des profils spectraux peut également être observée sur les zones asymptomatiques lorsque l'effet de la variété est pris en compte, suggérant des variations liées à la typologie de contamination. Toutefois, ces variations restent insuffisantes pour permettre une classification fiable des maladies dans ces zones.

Ces résultats démontrent le potentiel de la spectroscopie NIR pour la détection et la caractérisation des maladies de stockage de la pomme. Cela donne les perspectives pour le développement d'outils complémentaires pour le tri avec un diagnostic non destructif sur la ligne de calibration des fruits.

Références

[1] Gong D, Bi Y, Jiang H, et al. A comparison of postharvest physiology, quality and volatile compounds of 'Fuji' and 'Delicious' apples inoculated with Penicillium expansum. Postharvest Biology and Technology 2019;150:95–104. https://doi.org/10.1016/j.postharvbio.2018.12.018.

[2] Tahir II, Johansson E, Olsson ME. Improvement of Apple Quality and Storability by a Combination of Heat Treatment and Controlled Atmosphere Storage. Horts 2009;44:1648–54. https://doi.org/10.21273/HORTSCI.44.6.1648.

[3] Nicolaï BM, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 2007;46:99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024.

[4] Huang Y, Lu R, Chen K. Detection of internal defect of apples by a multichannel Vis/NIR spectroscopic system. Postharvest Biology and Technology 2020;161:111065. https://doi.org/10.1016/j.postharvbio.2019.111065.

[5] Bleasdale AJ, Whyatt JD. Classifying early apple scab infections in multispectral imagery using convolutional neural networks. Artificial Intelligence in Agriculture 2025;15:39–51. https://doi.org/10.1016/j.aiia.2024.10.001.

[6] Walsh KB, Blasco J, Zude-Sasse M, et al. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. Postharvest Biology and Technology 2020;168:111246. https://doi.org/10.1016/j.postharvbio.2020.111246.

Appui de la Spectroscopie Proche Infrarouge pour l'amélioration de l'identification et la discrimination des espèces d'arbres sur pied en forêt amazonienne

Max Hildebrandt [1], Daniela Florez Parra [2], Romain Lehnebach [2], Giacomo Sellan [2], Julie Bossu [3], Gilles Chaix [4]

1 Université de Montpellier, Master 2 BioGeT, AgroParisTech
2 UMR EcoFoG , CIRAD Guyane
3 UMR EcoFoG , CNRS Guyane
4 CIRAD - UMR AGAP Institut, Montpellier, France

Email : max.hildebrandt@etu.umontpellier.fr, daniela.florez@cirad.fr, romain.lehnebach@cirad.fr, giacomo.sellan@cirad.fr, julie.bossu@cnrs.fr, gilles.chaix@cirad.fr

***Mots-clefs** : Spectroscopie proche infrarouge, Arbre, Amazonie, Identification, botanique*

Introduction

L'identification botanique en milieu tropical peut s'avérer très complexe en regard de la forte diversité de ces écosystèmes et de la forte ressemblance morphologique des espèces proches sur le plan phylogénétique. Les difficultés d'identification entravent ainsi la compréhension du fonctionnement des forêts tropicales tout comme la valorisation des produits forestiers (Montagnini et al., 2005). Dans le cadre du projet NIRVANA (LabEx CEBA), ce travail évalue la capacité du NIRS à discriminer 15 espèces de Guyane française appartenant à trois familles à partir de quatre types de tissus végétatifs collectés. Deux questions principales sont adressées : (1) le NIRS permet-il de discriminer efficacement des espèces phylogénétiquement proches dans un contexte multi-familles ? (2) Quels tissus fournissent l'information spectrale la plus discriminante pour une identification terrain ? Pour y répondre, un pipeline chimiométrique complet a été développé.

Matériels et Méthodes

- Sites d'étude et méthode d'échantillonnage

Les échantillons ont été récoltés sur la station expérimentale de Paracou (Cirad), dans neuf parcelles permanentes du dispositif d'inventaire à long terme GuyaFor. Des échantillons complémentaires proviennent de la station des Nouragues (CNRS). Quatre types de tissus végétatifs ont été collectés par individu : feuilles fraîches (F), feuilles sèches (S), écorce externe (EE) et écorce interne (EI). Au total, 470 individus ont été mesurés, représentant environ 14 700 spectres. Les échantillons de référence (REF) sont issus de l'herbier IRD de Cayenne et se composent uniquement de feuilles sèches. Les acquisitions spectrales ont été réalisées à l'aide d'un spectromètre portable ASD LabSpec 5000 avec une sonde de contact de 6 mm de diamètre. Les échantillons ont été placés sur un fond Vantablack afin de minimiser les pertes de lumière. Un étalonnage sur spectralon blanc a été effectué toutes les 20 minutes pour corriger les dérives instrumentales. Cinq mesures ont été réalisées par feuille sur la face adaxiale sur trois feuilles par arbre, et trois mesures sur chaque face de l'écorce, pour un total de 36 mesures spectrales par individu.

- Traitement et analyse des données

L'analyse se limite aux régions 1000–1750 nm et 1850–2350 nm. La détection des spectres aberrants a été réalisée en deux passes successives. Quatorze combinaisons de prétraitements spectraux ont ensuite été évaluées de manière systématique pour chaque combinaison famille × tissu. Le prétraitement optimal a été

sélectionné sur la base de l'accuracy de classification en LOO (Leave One Out) et du coefficient de silhouette. L'ensemble des analyses a été conduit en RStudio (v4.5.1). Une analyse exploratoire préalable a été réalisée par famille et par tissu, combinant ACP, analyse inter-classes (BCA) et LDA. Trois méthodes discriminantes ont été comparées : LDA sur scores ACP, PLS-DA (validation M-fold 5×10, critère max.dist) et Random Forest (ntree = 1000, mtry = $\sqrt{p}$). La discrimination a été conduite à deux niveaux hiérarchiques, inter-familles puis intra-famille avec une progression des données d'entraînement intégrant progressivement les échantillons de référence herbier aux données de terrain. Les modèles finaux ont été entraînés sur la totalité des données REF (herbiers) combinée à 70% des données arbres TRN (terrain), et évalués sur deux jeux indépendants distincts (jeu test et jeu indépendant, 20% et 10% des arbres TRN respectivement), jamais utilisés lors de l'entraînement. La significativité des modèles a été vérifiée par tests de permutation (199 permutations, étiquettes Y_train permutées aléatoirement).

Résultats

- Qualité des données et analyses exploratoires

L'évaluation systématique des 14 combinaisons de prétraitements révèle que la combinaison detrend+SNV est optimale pour les feuilles sèches de Lecythidaceae et Burseraceae, tandis que savgol2 est retenue pour Lauraceae. Les performances LDA en LOO par arbre sur les données prétraitées varient de 54 à 95 % selon la famille et le tissu, avec une hiérarchie constante S > EI > F > EE. L'analyse ACP-BCA-LDA confirme une structuration spectrale nette entre les trois familles sur le tissu S (93.9 %, $\kappa = 0.904$), qui s'atténue sur F (85.0 %) et EE (73.4 %). Au niveau espèce, les Burseraceae (89.8 %) et les Lauraceae (89.5 %) sur S sont nettement plus discriminables que les Lecythidaceae (77.0 %), où la confusion entre *E. coriacea* et *E. sagotiana* constitue la principale source d'erreur sur tous les tissus. La comparaison des échantillons REF et TRN révèle que ces deux espèces sont parfaitement discriminées dans les herbiers (100% et 98.7%) mais présentent des performances nettement dégradées sur le terrain (74.2% et 60.2%), suggérant qu'elles sont fortement confondues sur le terrain et que les conditions de collecte masquent partiellement le signal discriminant de ces espèces.

- Modèles discriminants et performances finales

La PLS-DA inter-familles atteint 96.8% sur S en validation M-fold. Les modèles RF finaux entraînés sur REF + 70% TRN atteignent, sur le jeu indépendant, 95.6% ($\kappa = 0.939$) pour les Burseraceae, 83.5% ($\kappa = 0.777$) pour les Lauraceae et 67.2% ($\kappa = 0.578$) pour les Lecythidaceae. La PLS-DA se rapproche du RF uniquement pour les Lecythidaceae (62.9% vs 67.2% sur le jeu indépendant, mais 85.1% vs 95.4% sur le jeu test). Les tests de permutation confirment la significativité statistique des trois modèles ($p = 0$, 199 permutations). L'analyse des seuils de confiance révèle une hiérarchie opérationnelle claire : les Lauraceae sont directement déployable sans filtre avec 98.7% d'accuracy à 50% de seuil sur 99.2% des spectres acceptés, les Burseraceae atteignent 91.3% avec un seuil de 55%, tandis que les Lecythidaceae atteignent 97.3% dès 50% de seuil sur 92.7% des spectres acceptés, les 7.3% rejetés correspondant quasi-exclusivement aux confusions *coriacea*/*sagotiana*.

Conclusion

Cette étude démontre que la spectroscopie proche infrarouge permet la discrimination automatisée d'espèces forestières tropicales guyanaises. 10 espèces sur 15 atteignent des niveaux opérationnels sur jeu indépendant et tous les modèles étant hautement significatifs ($p < 0.005$). Le tissu foliaire sec est le support le plus discriminant pour les trois familles, la famille des Lauraceae étant directement déployable sans filtre de confiance. Les confusions persistantes entre espèces congénères reflètent un chevauchement spectral

géométriquement démontré ratios inter/intra inférieurs à 1 et partage strict des longueurs d'onde discriminantes entre paires problématiques et donc, non un défaut algorithmique. La comparaison des profils VIP herbier/terrain révèle cependant que certaines de ces limites ne sont pas biologiquement irréductibles : la séparation *E. coriacea/ E. sagotiana* existe dans le signal herbier à 87.6% mais est noyée par la variabilité du terrain. C'est une limite de standardisation du protocole de collecte terrain et de fiabilité des identifications botaniques initiales, pas une limite biologique, et elle constitue le levier prioritaire pour améliorer les performances du modèle au-delà des plafonds actuels.

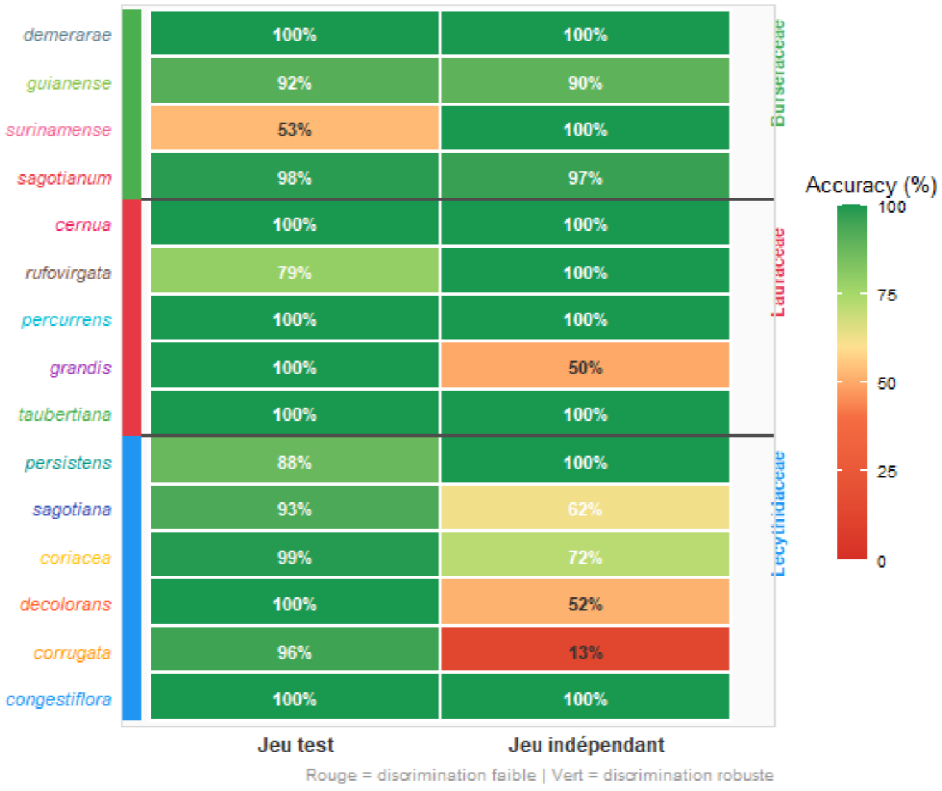


Figure 1 : Comparaison des performances de classification par espèce entre le jeu test (arbres jamais vus pendant l'entraînement, utilisés pour la sélection du modèle) et le jeu indépendant (arbres jamais vus et n'ayant influencé aucune décision méthodologique). Modèle Random Forest final (REF + 70% TRN, Split B), tissu S.

References

MONTAGNINI, Florencia et JORDAN, Carl F. *Tropical forest ecology: the basis for conservation and management*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2005.

Comment réduire la dépendance des réseaux de neurones aux valeurs de référence ? Les auto-encodeurs masqués, un cas appliqué d'apprentissage auto-supervisé

Ivy Tumoine[1,2,3,4], M. Metz[2,3,4,5], F. Abdelghafour[1,3,4], R. Bendoula[1,3,4], D. Esteve[2], J. M. Roger[1,3,4]

[1] UMR ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France
[2] Pellenc ST, Applied Research Group, Pertuis, France
[3] LabCom Aioly, Artificial Intelligence and Optics Laboratory, Montpellier, France
[4] ChemHouse, Research Group, Montpellier, France
[5] IMBE, Aix-Marseille University, UMR CNRS IRD Avignon University, Marseille, France

Email : i.tumoine@pellencst.fr

***Mots-clefs** : apprentissage profond, apprentissage auto-supervisé, masquage*

Les modèles d'apprentissage profond constituent une solution prometteuse pour la modélisation des relations complexes entre les spectres et les valeurs de référence. Pour atteindre des performances élevées, ils nécessitent de grandes quantités de données annotées, c'est-à-dire des échantillons avec des valeurs de référence. Or, ces données sont rares car l'obtention des valeurs de référence implique souvent des procédures de laboratoire coûteuses et chronophages. Réduire cette dépendance aux données annotées représente donc un défi majeur pour les applications spectroscopiques.

L'apprentissage auto-supervisé (SSL) offre une solution prometteuse en formulant des tâches prétextes qui tirent une supervision directement des données spectrales pour pré-entraîner des modèles d'apprentissage profond. Parmi les approches SSL, les auto-encodeurs masqués (MAE) [1] ont suscité un intérêt particulier pour leur simplicité et leurs performances. MAE divise le spectre d'entrée en patchs, masque un sous-ensemble et entraîne le modèle à reconstruire les patchs manquants.

Bien que MAE ait été appliqué aux données spectroscopiques [2], [3], son potentiel pour réduire les besoins d'annotation en spectroscopie proche infrarouge (SPIR) reste peu exploré. Dans ce travail, un pré-entraînement MAE simple est évalué sur deux tâches : la classification de variétés de blé à partir de spectres PIR [4] et la prédiction de la teneur en carbone organique du sol à partir de spectres Vis-PIR [5]. L'influence des principaux paramètres du MAE est étudiée. Les résultats soulignent le potentiel du pré-entraînement MAE pour les données PIR (Figure 1) et fournissent des pistes pour l'amélioration des stratégies de masquage spectral et le développement de modèles de fondation pour les données PIR.

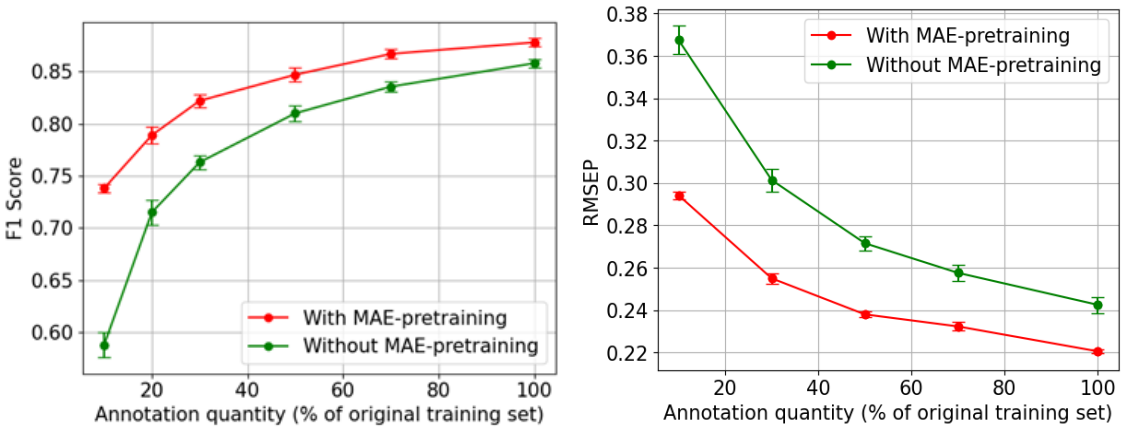


Figure 1 : Performance du modèle en fonction de la quantité de données annotées, avec et sans pré-entraînement MAE, pour la classification des variétés de blé (à gauche) et la prédiction de la teneur en carbone organique du sol (à droite).

References

[1] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, et R. Girshick., *arXiv, 2021,* doi: 10.48550/arXiv.2111.06377.

[2] P. Ren, R.-G. Zhou, et Y. Li, *Expert Syst. Appl.,* 2025, vol. 292, p. 128576, doi: 10.1016/j.eswa.2025.128576.

[3] M. Wan *et al.*, *Comput. Electron. Agric.*, 2023, vol. 215, p. 108427, doi: 10.1016/j.compag.2023.108427.

[4] L. Zhou *et al.*, *Front. Plant Sci.*, 2020, vol. 11, doi: 10.3389/fpls.2020.575810.

[5] J. L. Safanelli *et al.*, 2023, doi: 10.1101/2023.12.16.572011.

Développement d'un réseau de micro-spectromètres NIR pour la discrimination d'espèces forestières malgaches

Randriambinintsoa Tiavina[1,2,3], Ramananantoandro Tahiana[1,2], Chaix Gilles[3,4,5]

[1]*Université d'Antananarivo, Ecole Supérieure des Sciences Agronomiques, Mention Foresterie et Environnement, Antananarivo 101, Madagascar*
[2]*Université d'Antananarivo, Ecole Doctorale Gestion des Ressources Naturelles et Développement*
[3]*ChemHouse, Research Group, Montpellier, France*
[4]*CIRAD - UMR AGAP Institut, Montpellier, France*
[5]*UMR AGAP Institut, Univ. Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*

Mail : randriambinintsoat@gmail.com

***Mots-clefs** : bois précieux ; discrimination ; Madagascar ; portable ; spectrométrie proche infrarouge ; terrain ; transfert d'étalonnage*

Introduction

La préservation des forêts malgaches, confrontées à une exploitation illégale intense du bois, requiert des outils rapides et fiables d'identification des essences. La spectroscopie dans le proche infrarouge (SPIR ou NIR), associée à des modèles chimiométriques tels que la PLS-DA, a démontré son efficacité pour discriminer les espèces forestières tropicales, notamment du genre *Dalbergia*, avec des taux de classification souvent supérieurs à 90 % (Pan et al., 2022 ; Chen et al., 2025).

Cependant, le passage du laboratoire au terrain reste complexe. Les variations environnementales (humidité, température), la préparation des surfaces et surtout les différences instrumentales entre spectromètres de laboratoire (Bruker MPA II) et portables (Innospectra NIR-SG1) limitent la transférabilité des modèles (Lazarescu et al., 2017 ; Xue et al., 2022).

Cette étude vise à surmonter ces obstacles par le transfert d'étalonnage entre appareils. En utilisant des approches de concaténation spectrale (UPDATE et Double UPDATE), elle permet de valoriser les bases de données acquises en laboratoire et de garantir des prédictions fiables sur les spectromètres portables.

Matériel et méthodes

Appareils de mesure :
Les mesures ont été réalisées avec un spectromètre de laboratoire Bruker MPA II (Bruker Optik GmbH, Allemagne) avec une gamme spectrale entre 850–2500 nm et de quatre spectromètres portables Innospectra NIR-SG1 (InnoSpectra, Taiwan ; 900–1700 nm, technologie DLP®).

Échantillons et données spectrales :
Deux cents micro-carottes de bois (5 mm de diamètre) ont été prélevées à 1,3 m de hauteur sur quatre espèces de *Dalbergia* (bois précieux de Madagascar). Les échantillons ont été répartis en trois groupes : étalonnage (138 carottes, Bruker), standards communs (32 carottes) et test (30 carottes), mesurés sur le Bruker et les quatre Innospectra (Tableau 1).

Méthodologie d'étalonnage et de transfert d'étalonnage :

Le Bruker MPA II a servi de spectromètre source et les Innospectra SG1 de spectromètres cibles. Les spectres Bruker ont été transformés pour correspondre à la gamme et à la résolution des Innospectra (Figure 1). Les modèles de discrimination ont été développés par PLS-DA (algorithme NIPALS) avec le package rchemo (Brandolini-Bunlon et al., 2023) sous R 4.5.0 (R Core Team, 2024 ; RStudio). Le nombre optimal de variables discriminantes et le meilleur prétraitement (brute, DT, SNV, SG, SG1, SG2 et combinaison 2 à 2) ont été sélectionnés par validation croisée (4 blocs, 20 répétitions) en minimisant l'erreur de classification.
Le transfert d'étalonnage a été réalisé par les méthodes UPDATE et Double UPDATE (Giordanengo et al., 2008 ; Roger, 2010), répétées pour chacun des quatre spectromètres cibles. Cinq erreurs globales de classification ont été évaluées (ERR1 à ERR5, Figure 1).

Tab. 1 : Effectifs des spectres par espèces, par lots et par spectromètres

Espèces Répartition en Lot	Spectromètre de mesure	*Dalbergia greveana*	*Dalbergia monticola*	*Dalbergia orientalis*	*Dalbergia purpurascens*
Etalonnage (138 carottes)	Bruker	108	204	282	234
Standards (32 carottes)	Bruker	36	48	48	60
	Innospectra 01 à 09	324 (36 * 9)	432 (48 * 9)	432 (48 *9)	540 (60 * 9)
Test (30 carottes)	Bruker	30	48	48	54
	Innospectra 01 à 09	270 (30*9)	432 (48 * 9)	432 (48 * 9)	486 (54 *9)

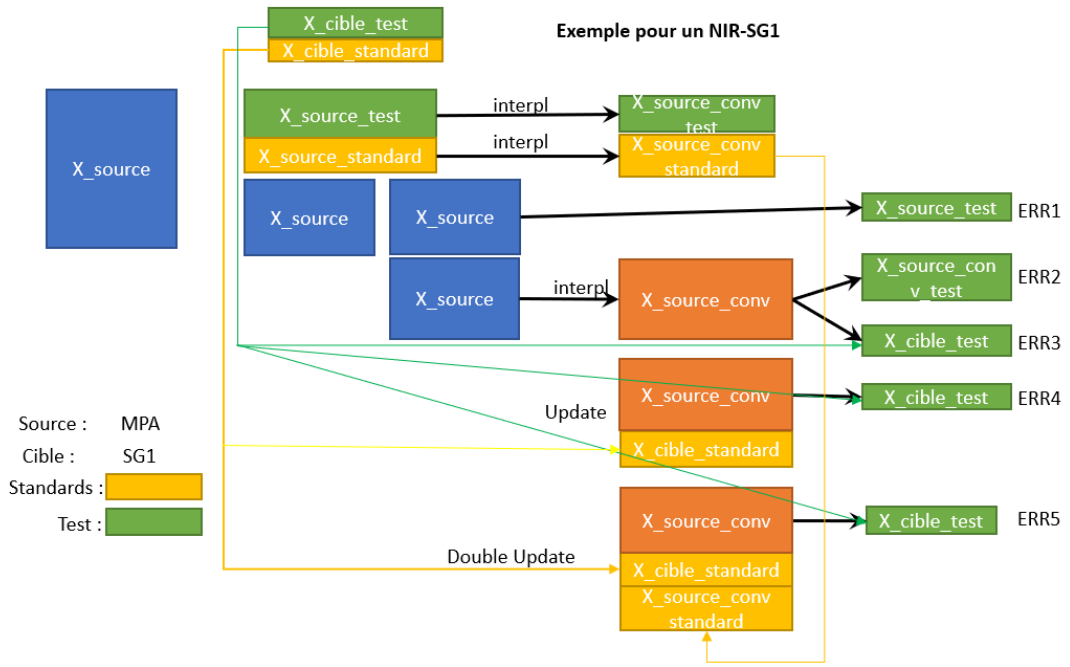


***Fig.1** : Résumé méthodologique pour le processus de transfert d'étalonnage entre Bruker et Innospectra*

Résultats et discussion

Le transfert d'étalonnage par concaténation des données spectrales s'est révélé efficace, permettant de transférer les modèles PLS-DA du Bruker MPA II vers les spectromètres portables Innospectra NIR-SG1 avec des performances proches de celles obtenues sur l'appareil source pour les quatre espèces de Dalbergia.

Tab.2 : *Erreur globale des modèles de discrimination en fonction des données d'étalonnage et de tests*

Tests Etalonnage	Pret / VD	X_source_test_1000-2500	X_source_conv_test	X_cible_1_test	X_cible_2_test	X_cible_3_test	X_cible_4_test
Gamme : 1000-2500		N=180 (30/48/48/54)					
X_source N=828 108/204/282/234	SG1 / 18	ERR1 42,2					
Gamme : 900-1700			ERR2	ERR3	ERR3	ERR3	ERR3
X_source N=828 108/204/282/234	SG2SNV / 12		45,2	52,2	51,6	54.4	55.6
Gamme : 900-1700				ERR4	ERR4	ERR4	ERR4
Upt1 (X_source + X_cible) N=1020 144/252/330/294	SG2SNV / 19			37.2	58.3	41.1	62.2
Gamme : 900-1700				ERR5	ERR5	ERR5	
Upt2 (X_source + X_cible + X_source_standard) N=1212 180/300/378/354	SG0 / 25			41.6	54.4	38.3	53.3

Ces résultats ouvrent la voie à l'enrichissement des modèles avec d'autres espèces de *Dalbergia* et à leur extension aux autres bois précieux, notamment le genre *Diospyros*. Cependant, la PLS-DA présente une limite importante : assigne systématiquement tout échantillon à une classe connue, risquant de fausses identifications sur bois ordinaires ou autres matériaux (plastiques, métaux, bois communs). L'intégration de méthodes pour la détection des « out-of-distribution » (OOD) comme One-Class SVM, SIMCA est donc nécessaire pour améliorer la robustesse des modèles pour les identifications directement sur le terrain. Au final, ces avancées pourraient renforcer les outils de contrôle terrain et contribuer à une meilleure gestion

durable des bois précieux malgaches. Ces avancées pourraient faciliter le déploiement d'un réseau de µ-spectromètres NIR pour renforcer le contrôle du commerce du bois précieux malgache, tant par les autorités forestières (administration forestière lors des contrôles ou en sortie de forêt, sur les routes nationales) que par les services douaniers (au niveau des ports).

References

Brandolini-Bunlon, M., Jallais, B., Roger, J. M., & Lesnoff, M. rchemo: Dimension Reduction, Regression and Discrimination for Chemometrics R package version 0.1-2. 2023, https://cran.r-project.org/package=rchemo

Chen, Z., Xue, X., Wu, H., Gao, H., Wang, G., Ni, G., & Cao, T. Visible/near-infrared hyperspectral imaging combined with machine learning for identification of ten Dalbergia species. Frontiers in Plant Science, 15, 2024, Article 1413215. https://doi.org/10.3389/fpls.2024.1413215

Lazarescu, C., Hansmann, H., & Rautkari, L. Wood species identification by near-infrared spectroscopy. International Wood Products Journal, 2017. 32–35. https://doi.org/10.1080/20426445.2016.1242270

Pan, X., Qiu, J., & Yang, Z. Identification of five similar Cinnamomum wood species using portable near-infrared spectroscopy, 2022, 37, 28–34. https://doi.org/10.56530/spectroscopy.zg7089n4

Xue, X., Chen, Z., Wu, H., Gao, H., Nie, J., & Li, X. Identification of eight Pterocarpus species and two Dalbergia species using visible/near-infrared (Vis/NIR) hyperspectral imaging (HSI). Forests, Article 1259. 2023, https://doi.org/10.3390/f14061259

Étude de l'adultération des épices en utilisant la spectroscopie visible et proche infrarouge combinée à la chimiométrie et l'apprentissage automatique.

Mohamed-Amine Antar [1], Youssef Tmimi [1], Fouad Fethi [1], Mounim Chikri [1]

[1] *Laboratoire de Physique de la Matière et de Rayonnements (LPMR), Faculté des Sciences, Université Mohammed Premier, Oujda, Maroc.*

Email : mohamed-amine.antar.m24@ump.ac.ma

***Mots-clefs** : Spectroscopie visible/proche infrarouge, adultération, apprentissage automatique, poivre noir, contrôle qualité.*

Les épices occupent une place fondamentale sur le marché agroalimentaire mondial et constituent un élément essentiel du patrimoine culinaire, notamment dans la riche gastronomie marocaine. Cependant, en raison de leur forte demande et de leur valeur économique, ces produits sont devenus extrêmement vulnérables à diverses pratiques d'adultération frauduleuses à des fins purement commerciales. Face à cet enjeu de contrôle qualité, cette étude s'intéresse au développement de méthodes de détection rapides et non destructives pour évaluer l'authenticité des épices couramment consommées, telles que le gingembre [1], le paprika [2] et le poivre noir.

Cette étude porte sur le cas du poivre noir et vise à quantifier son taux d'adultération par le gui blanc, en utilisant la spectroscopie visible (Vis) et proche infrarouge (PIR) comme méthode rapide et non destructive [3]. Afin d'optimiser l'exploitation des données spectrales collectées, une démarche chimiométrique a été déployée, incluant l'application de diverses techniques de prétraitement aux signaux bruts pour éliminer le bruit [4], suivies par l'Analyse en Composantes Principales (ACP) pour réduire la dimensionnalité et extraire l'information pertinente [5].

La phase de quantification a été modélisée en s'appuyant sur des algorithmes d'apprentissage automatique, en comparant les performances de la Régression à Vecteurs de Support (SVR) et des Réseaux de Neurones Artificiels (RNA). L'évaluation des performances de ces algorithmes s'est appuyée sur le coefficient de détermination (R^2) et l'erreur quadratique moyenne (RMSE). Les résultats démontrent que les RNA, couplés aux spectres PIR, fournissent le modèle prédictif le plus robuste, affichant des métriques optimales avec un Rp^2=0,9734 et un RMSEP=1,3582.

Références

[1] M. Chikri, L. Srata, S. Farres, Y. Tmimi, and F. Fethi, "The development of a green analytical method to monitor adulteration in ginger using visible and near-infrared spectroscopy combined with chemometric tools," *Moroccan J. Chem.*, vol. 13, no. 1, pp. 122–132, 2025, doi: 10.48317/IMIST.PRSM/morjchem-v13i1.46012.

[2] Y. Tmimi, M. Chikri, and F. Fethi, "Rapid quantification of paprika adulteration by food dye using visible and near-infrared spectroscopy combined with machine learning," *Food Humanit.*, vol. 5, Dec. 2025, doi: 10.1016/j.foohum.2025.100834.

[3] S. Farres, L. Srata, F. Fethi, and A. Kadaoui, "Argan oil authentication using visible/near infrared spectroscopy combined to chemometrics tools," *Vib. Spectrosc.*, vol. 102, no. September 2018, pp. 79–84, 2019, doi: 10.1016/j.vibspec.2019.04.003.

[4] L. Srata, M. Chikri, S. Farres, I. Hamdani, Y. Tmimi, and F. Fethi, "A comparative study of discrimination and parameter identification of oil types using near-infrared spectroscopy, fourier transform infrared spectroscopy, and laser-induced fluorescence spectroscopy combined with chemometrics tools," *Interactions*, vol. 246, no. 1, Dec. 2025, doi: 10.1007/s10751-024-02239-8.

[5] A. B. S. de Lima, A. S. Batista, J. C. de Jesus, J. de J. Silva, A. C. M. de Araújo, and L. S. Santos, "Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling," *Food Control*, vol. 107, Jan. 2020, doi: 10.1016/j.foodcont.2019.106802.

Posters

Mesures *in situ* par SPIR de la qualité de chênes pédonculés abroutis ou non par le chevreuil

A. Bled[1,5], D. Bastianelli[2,3], A. Gnanhoui[2,3], L. Bonnal[2.3], V. Boulanger[4], C. Collet[1], S. Saïd[5]

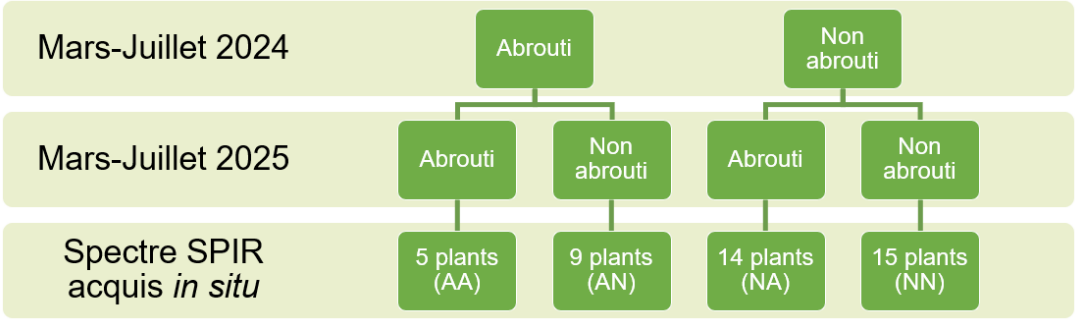
[1] *Université de Lorraine, AgroParisTech, INRAE, UMR Silva, 54000, Nancy, France.*
[2] *CIRAD, UMR SELMET, F-34398 Montpellier, France.*
[3] *SELMET, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France.*
[4] *Office National des Forêts, Département Recherche, Développement et Innovation, 77300, Fontainebleau, France.*
[5] *Office Français de la Biodiversité, Direction Recherche et Appui Scientifique, 01330, « Montfort » Birieux, France.*

Email : sonia.said@ofb.gouv.fr

***Mots-clefs** : Capreolus capreolus, Quercus robur, NIRS, abroutissement, qualité nutritive*

La pression d'herbivorie causée par les ongulés sauvages représente un enjeu pour la régénération forestière dans de nombreux pays de l'hémisphère nord. Un abroutissement intense tend à réduire la croissance des semis et à entraîner une mortalité accrue. L'attractivité des plants pour les ongulés dépend notamment de leur qualité nutritive. Cette qualité nutritive des plants peut elle-même évoluer en fonction de l'intensité et de la fréquence d'abroutissement. Cependant, Les méthodes actuelles permettant d'estimer la qualité nutritive des plants sont généralement destructives et ne permettent pas de faire d'études longitudinales.

L'étude a été conduite sur le site du Territoire d'Etudes et d'Expérimentations (TEE) de l'OFB à Trois-Fontaines (Grand Est, France). À l'aide d'un spectromètre portable (VIAVI MicroNIR ; spectres 125 L.O., gamme 900-1680 nm) , nous avons évalué *in situ* et de manière non destructive la qualité nutritive de jeunes chênes pédonculés (*Quercus robur*) après deux ans d'exposition à la pression d'abroutissement du chevreuil (*Capreolus capreolus*). Les chênes pouvaient avoir été abroutis (A) ou non abroutis (N) en année 1 et en année 2, générant 4 modalités (NN, AN, NA, AA). La collecte des spectres sur le terrain s'est révélée rapide (quelques minutes par plant) et techniquement robuste sous réserve d'une connexion stable entre le spectromètre et la tablette d'acquisition.



A partir de ces spectres de feuilles, nous avons prédit les paramètres de qualité nutritive : Teneur en protéines (MAT, %MS), Teneur en fibres (NDF, %MS), Digestibilité (Solubilité de la matière organique SMO, %MO). Les étalonnages utilisés étaient préliminaires et basés sur une base d'analyses de référence de plantes ligneuses.

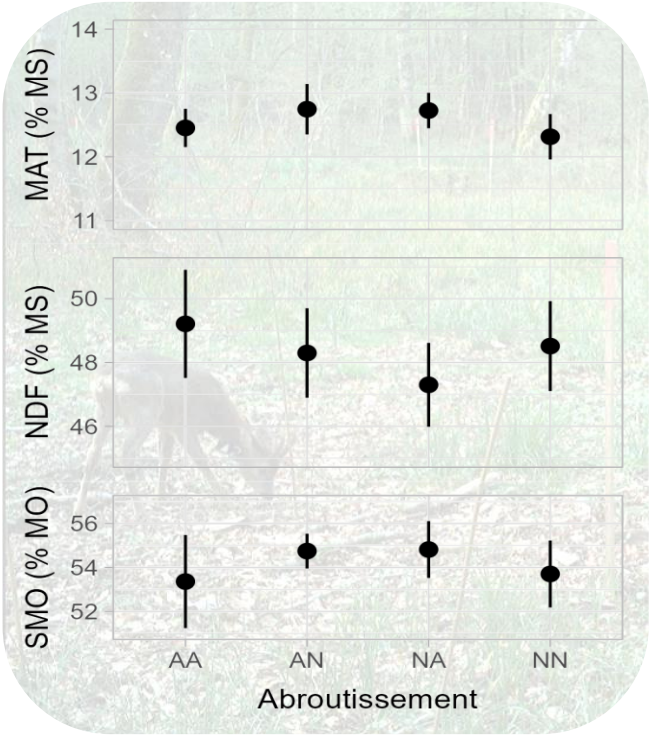


***Figure 1**. Composition chimique prédite des feuilles de chêne selon les modalités d'abroutissement*

Nous n'avons pas constaté de différence significative de la teneur en protéines, en fibres, ni de la digestibilité des feuilles de chêne en fonction de l'abroutissement. Cependant, le faible nombre de plants abroutis et la précision encore limitée des étalonnages NIR, construits à partir d'un faible nombre d'échantillons de référence, ont réduit notre capacité à détecter d'éventuelles différences entre modalités. Il est donc nécessaire de poursuivre l'analyse avec un plus grand nombre de plants.

Au plan méthodologique, l'étude a confirmé que les mesures de spectre NIR directes sur le terrain permettent d'assurer un suivi longitudinal de jeunes plants sans perturber leur développement.

Early prediction of tuber yield in yam (*Dioscorea* spp.) using NIR spectra from pre-planting tubers and mature leaves

Emeline Legros[1,2], Lévy Laurent[1], Christophe Perrot[1], Erick Malédon[1], Marie-Claire Gravillon[1], Elie Nudol[1], Saskia Sergeant[1], Brice Rose-Antoinette[1], Steeve Joseph[1], Hanâ Chaïr[3,4], Komivi Dossa[1,3*]

[1] *CIRAD, UMR AGAP Institut, 97170 Petit Bourg, Guadeloupe, France*
[2] *Institut Polytechnique UniLaSalle, 60000 Beauvais, France*
[3] *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France*
[4] *CIRAD, UMR AGAP Institut, F-34398 Montpellier, France*

**Email : komivi.dossa@cirad.fr*

***Keywords** : NIRS, Phenomic prediction, Yam breeding, Genetic gain*

INTRODUCTION

Yam (*Dioscorea* spp.) is a staple crop for over 500 million people [1], yet breeding progress is slowed by vegetative propagation, dioecy, high heterozygosity, occasional polyploidy and long cycles, which considerably slow genetic progress [2]. Phenomic prediction using high-dimensional traits such as Near-Infrared Spectroscopy (NIRS) offers a rapid, non-destructive and low-cost method to predict complex agronomic traits early in the cycle, potentially saving months or an entire selection cycle [3, 4, 5]. This study tests whether NIRS, acquired on pre-planting tubers or on 3-month-old mature leaves, can reliably predict tuber yield across genotypes, years and sites, and which sampling and modelling choices maximize operational value in breeding.

MATERIAL AND METHODS

Plant material and trials: 204 yam genotypes from CIRAD's working collection (multiple *Dioscorea* species) [6] were evaluated over 3 campaigns (2023/24, 2024/25, 2025/26) at 2 experimental sites in Guadeloupe to capture multi-environment variation.

Spectral acquisition and preprocessing: Spectra were measured with the QualitySpec® TREK ASD across 350–2500 nm on 2 organs (tubers before planting and leaves at 3 months) and 2 sample preparations (fresh material and dried/ground flour) (Figure 1). After testing 15 preprocessing methods, Standard Normal Variate (SNV) followed by a Savitzky–Golay 2nd-derivative (SG2) was selected.

Prediction models and validation: 4 prediction approaches were compared, PLS (ncomp=10), pedigree BLUP (pBLUP), RKHS (nIter=5000, burnIn=1000) and Random Forest (ntree=300), with performance assessed by repeated cross-validation (25 repeats of 5-fold). Three realistic breeding scenarios were tested: CV1 (early selection of new genotypes), CV2 (prediction across related environments), and CV0 (prediction in new environments) using tuber yield (t/ha) as the target trait.

RESULTS

Variance decomposition and heritability: Spectra from fresh tubers showed the largest genetic contribution (31%), whereas flour spectra (leaves and tubers) were dominated by genotype × environment interaction (~70%) with low residual variance (6–8%); fresh samples, especially fresh leaves, had higher residual variance (≈57%), indicating greater experimental or measurement noise. ANOVA showed a significant year effect on tuber yield but no significant site effect, with high overall heritability ($H^2 \approx 0.787$).

Organ and sample-type comparison: Prediction accuracies were similar between leaf and tuber spectra; fresh material tended to yield slightly better predictions than flour. Across models, pBLUP gave the best average predictive performance (Figure 1a).

Prediction in breeding scenarios: For early selection of new genotypes (CV1), pBLUP achieved correlations around 0.51. Prediction across environments (CV2) showed variable accuracies between years (0.28–0.59). Prediction in entirely new environments (CV0) was the most challenging: accuracies dropped markedly and became more variable, notably when extrapolating across years (0.17–0.22). Predictions within the same campaign remained reasonable (0.37–0.58), emphasizing the influence of year-to-year environmental variation on model transferability (Figure 1b).

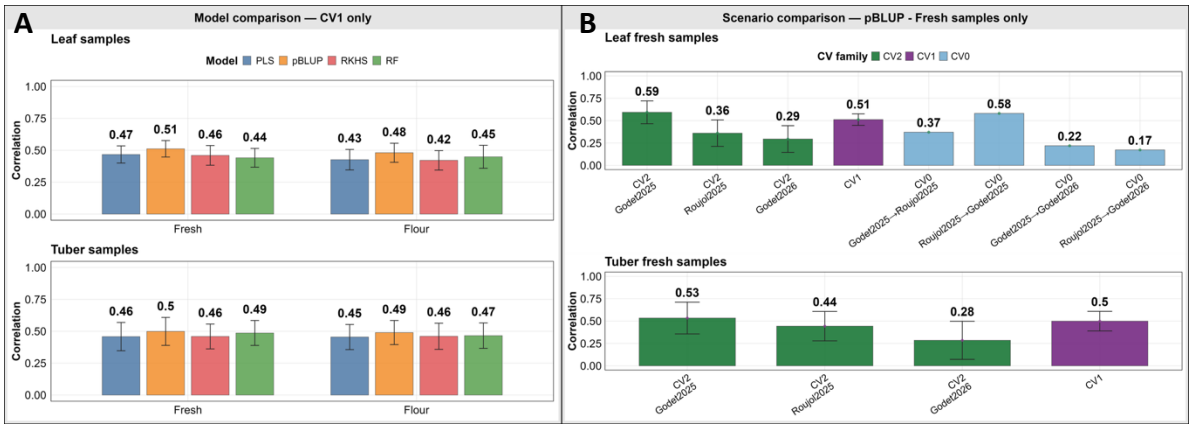


Figure 1: Predictive performance of phenomic models. a) *Comparison of predictive performances across sample types and statistical models*, **b)** *Comparison of breeding scenarios using the pBLUP model for fresh samples only*

CONCLUSION AND PERSPECTIVES

The results demonstrate that NIRS can provide actionable early predictions of tuber yield, particularly when using fresh samples (tubers or leaves) and pBLUP model, enabling earlier selection decisions that could shorten breeding cycles. The greater genetic signal in fresh tuber spectra and the slightly superior performance of fresh over flour samples indicate that minimal preprocessing (fresh sampling) may be both effective and more time-efficient for routine breeding workflows. Predictive performances remained satisfactory for early selection (CV1) and transfer across related environments (CV2) but decreased under new-environment scenarios (CV0), highlighting the major influence of environmental effects that still need to be integrated into prediction frameworks. Overall, these findings suggest that spectral phenotyping could become a powerful tool to accelerate yam breeding, provided that sample type, preparation methods, and prediction models are carefully optimized.

Future work will extend phenomic prediction to additional traits, including disease resistance, dry matter content, and taste quality, identify informative wavelength regions for trait-specific predictive models, and validate these approaches in hybrid populations derived from parental lines to ensure their applicability in real-world breeding and selection pipelines.

FUNDING

This study was supported by HORIZON Europe “ROTATES : Minor root and tuber crops fostering agrobiodiversity and ecosystem services ” project (number: 101181532) and European Union and

Guadeloupe Region (Programme Opérationnel FEDER – Guadeloupe - Conseil Régional 2021-2027) “TAESA” project (number: GPE005599)

REFERENCES

[1] Asiedu, R., & Sartie, A. (2010). Crops that feed the World 1. Yams. Food Security, 2(4), 305-315. https://doi.org/10.1007/s12571-010-0085-0

[2] Malédon, E., Nudol, E., Perrot, C., Gravillon, M.-C., Rivallan, R., Cornet, D., Chair, H., & Dossa, K. (2023). First Report of a Successful Development of Yam Hybrids (*Dioscorea alata* L.) from Lyophilized and Long-Term Stored Pollens. ResearchGate, 92(10). https://doi.org/10.32604/phyton.2023.042397

[3] Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science & Technology, 18(2), 72-83. https://doi.org/10.1016/j.tifs.2006.09.003

[4] Robert, P., Brault, C., Rincent, R., & Segura, V. (2022). Phenomic Selection : A New and Efficient Alternative to Genomic Selection (GS). In N. Ahmadi & J. Bartholomé (Éds.), Genomic Prediction of Complex Traits : Methods and Protocols (p. 397-420). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_14

[5] Zhu, X., Maurer, H. P., Jenz, M., Hahn, V., Ruckelshausen, A., Leiser, W. L., & Würschum, T. (2022). The performance of phenomic selection depends on the genetic architecture of the target trait. Theoretical and Applied Genetics, 135(2), 653-665. https://doi.org/10.1007/s00122-021-03997-7

[6] Dossa, K., Arnau, G., Malédon, E., Nudol, E., Gravillon, M.-C., Perrot, C., Laurent, L., Sergeant, S., Uneau, Y., Umber, M., Casi, D., Irep, J.-L., Hammouya, D., Andypain, S., Hubert, O., Chilin-Charles, Y., Louisor, J., Hery, M., Fouks, B., … Chaïr, H. (2026). YamHub as an international platform for yam research and breeding based in Guadeloupe. Nature Genetics, 1-3. https://doi.org/10.1038/s41588-026-02520-2

Jchemo: Chemometrics and machine learning on high-dimensional data with Julia

Matthieu Lesnoff [1,2,3]
[1] *SELMET, Univ Montpellier, CIRAD, INRAe, Institut Agro, Montpellier, France*
[2] *CIRAD, UMR SELMET, Montpellier, France*
[3] *ChemHouse Research Group, Montpellier, France*

Email : matthieu.lesnoff@cirad.fr

Key-words *: Chemometrics, Machine learning, Julia language, Toolbox*

Introduction

Julia (https://julialang.org) is a programming language designed for high computing performance. It is an open-source project made available under the MIT license. The language tries to tackle the "two-language problem". This problem refers to the fact that many scientific codes are prototyped in a flexible language (to test an idea quickly), but slow, and then have to be moved to a faster (e.g., C++) but less flexible language for practical applications. Julia allows both easily readable coding and fast computations. Works on Julia began in 2009. Julia's syntax is now considered stable, since version 1.0 in 2018 (actual version July 2026: 1.12.6), with many registered available packages and a very active users' forum (https://discourse.julialang.org). The proposed poster present Jchemo [1] (https://github.com/mlesnoff/Jchemo.jl), a Julia-written ***tool-box*** dedicated to chemometrics and machine learning in general.

(1) Why did I decide to switch from the language R to Julia in 2021 for my chemometrics works? Trying to run a PLSR (25 LVs) with n = 1e6 samples and p = 500 variables with my function crashed systematically my R working sessions (with a I9 Intel processor). With the same computer, and functions written in Julia, the computation only lasted 8 seconds. ***(2) Why did I choose Julia compared to Matlab?*** Since Julia is free.

Julia automatically compiles the functions to efficient native code via LLVM, and support multiple platforms (Windows, MacOs, Linux etc.). Julia uses ***multiple dispatch*** as a paradigm, making it easy to express many class and functional programming pattern. The official IDE recommended for Julia users is Visual Studio Code (https://code.visualstudio.com).

Material and methods

Jchemo was built initially around partial least squares regression (PLSR) and discrimination (PLSDA) methods and their non-linear extensions, in particular locally weighted PLS models (kNN-LWPLS-R & -DA; e.g., [2]). The package has then been expanded with many other methods of dimension reduction, regression, discrimination, and signal (e.g., spectra) preprocessing.

Why the name Jchemo? Since it is oriented towards chemometrics, in brief the use of biometrics for chemistry data. But most of the provided methods are generic and can be applied to other types of data. The package has two related projects: **JchemoData** (a container package of data sets used in the examples; https://github.com/mlesnoff/JchemoData.jl) and **JchemoDemo** (a pedagogical environment; https://github.com/mlesnoff/JchemoDemo).

Beside usual chemometrics methods (signal preprocessing, PCA, PLS etc.), multi-block methods are available for dimension reduction (e.g., MBPCA, ComDim, rCCA, etc.) and regression/discrimination

(MBPLS, ROSAPLS, SOPLS, etc.). Various ridge and sparse models are proposed, as many nonlinear models useful for analyzing heterogeneous data (kernels-, kNN-, Tree--based models, etc.). The syntax of Jchemo is very consistent between all the functions and implementation of methods can therefore be easily done by non-specialists of programming.

Results and discussion

Jchemo functions are organized between: ***transformation operators*** (e.g., PCA models), ***predictors*** (e.g., PLSR/PLSDA models), and utility functions.

Ad'hoc pipelines (chains of models) can also easily be built. In Jchemo, a pipeline is a chain of K models: *(a)* either a set of K transformers, or *(b)* a set of K – 1 transformers and a final predictor. The fitting of an ad'hoc pipeline is illustrated below. The example considers the "LWR" algorithm of Naes et al. [3] that consists in chaining two models (more than two models can be chained): a preliminary global PCA on the data and then a kNN locally weighted multiple linear regression (kNN-LWMLR) on the global PCA scores.

```
model1 = pcasvd(; nlv = 25)                          # transformation operator
model2 = lwmlr(; metric = :eucl, h = 2, k = 200)     # predictor
model = pip(model1, model2)  # final pipeline
fit!(mod, X, Y)
pred = predict(mod, Xnew).pred
```

Tuning functions. Efficient generic (i.e., the same for all models) functions allow to tune the models, by test-set validation or cross-validation. For the first, a grid-search for a gaussian KPLSR model can for instance be implemented by:

```
kern = [:krbf] ; gamma = [100, 1, .1, 0.001]
pars = mpar(kern = kern, gamma = gamma)  # the grid
nlv = 0:30
model = kplsr()
res = gridscore(model, Xcal, Ycal, Xval, Yval; score = rmsep, pars, nlv)
```

Jchemo is registered on the official Julia package repository (equivalent of the CRAN for R). It is an easy tool that can handle the usual needs in chemometrics, as well as for baseline studies or for research activities.

The development of the package is active. The package is regularly updated by new functions. For instance, sparse PCA and PLS functions, one-class-classification (OCC) functions, etc., have been recently added.

References

[1] M. Lesnoff. Jchemo: Chemometrics and machine learning on high-dimensional data with Julia. 2021, https://github.com/mlesnoff/Jchemo. UMR SELMET, Univ Montpellier, CIRAD, INRA, Institut Agro, Montpellier, France.

[2] M. Lesnoff, M. Metz, J.M. Roger. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. Journal of Chemometrics n/a, https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.3209.

[3] T. Naes, T. Isaksson, B. Kowalski. Locally weighted regression and scatter correction for near-infrared reflectance data. Analytical Chemistry 664–673, 1990. https://pubs.acs.org/doi/10.1021/ac00206a003.

PRO-PIX / ONE-PIX : une approche d'imagerie hyperspectrale mono-pixel pour la spectroscopie appliquée et la production d'indicateurs embarquée

Mathieu Ribes [1], Gaspard Russias [1], Thomas Lebrat [1] et Antoine Fournier [1]

[1] *Photonics open Projects, 4 rue Louis de Broglie - 22300 Lannion/France*

Email : mribes@photonics-open-projects.com, afournier@photonics-open-projects.com

***Key-words** : spectroscopie NIR/SPIR, imagerie hyperspectrale, acquisition compressée, mono-pixel, chimiométrie, apprentissage statistique, transfert de calibration, traitement embarqué, instrumentation sobre*

Les approches spectroscopiques dans le domaine VIS–NIR–SWIR occupent aujourd'hui une place centrale dans de nombreux champs d'application, allant de l'agronomie à l'environnement, de l'agro-industrie à l'observation de la Terre, en passant par le contrôle qualité et les sciences des matériaux. Elles permettent d'accéder à des informations physico-chimiques à partir de signatures spectrales, mais leur déploiement opérationnel reste contraint par les coûts instrumentaux, la gestion de volumes de données élevés et les enjeux de robustesse des modèles en conditions réelles.

Dans ce contexte, Photonics Open Projects (POP), spin-off du partenariat AgroPhotonique entre ARVALIS – Institut du Végétal et Photonics Bretagne, développe une approche d'imagerie hyperspectrale compressée reposant sur une architecture mono-pixel reconfigurable. Cette approche vise à transformer des spectromètres standards en dispositifs imageants, tout en permettant l'exécution de modèles chimiométriques embarqués à même de produire directement, au point de mesure, des estimations quantitatives et des indicateurs décisionnels en temps réel.

Le kit ONEPIX adresse à la fois l'enseignement, la recherche et le prototypage algorithmique en imagerie hyperspectrale, avec une approche ouverte et reproductible facilitant le développement et le partage de méthodes. Elle se décline également en une gamme PROPIX pour des applications de terrain et industrielles (phénotypage, tri, contrôle en ligne, intégration OEM), ainsi qu'en une offre EDU dédiée à l'intégration pédagogique et à l'appropriation des concepts avancés en spectroscopie et traitement des données. Un appui de Bureau d'Etude permet l'embarquement d'algorithmes de traitement avancés, incluant segmentation, apprentissage statistique et estimation spectro-chimiométrique.

Le système a été évalué dans différents contextes expérimentaux, illustrant sa capacité à produire des données exploitables en continu tout en réduisant fortement les volumes générés. Cette approche ouvre la voie à une instrumentation sobre, orientée extraction d'information plutôt que production massive de données, en cohérence avec les enjeux actuels de déploiement à grande échelle et de déploiement numérique raisonné.

Les travaux en cours portent sur la robustesse des modèles en conditions réelles, le transfert d'étalonnage entre instruments et contextes d'acquisition, ainsi que la prise en compte et la propagation des incertitudes associées aux estimations. Ces problématiques, au cœur des verrous actuels en spectroscopie appliquée et en apprentissage sur données spectrales, s'inscrivent pleinement dans les thématiques portées par la communauté HélioSPIR, en particulier autour de la généricité des modèles, de leur transférabilité et de leur intégration embarquée.

POP propose ainsi une instrumentation modulaire, ouverte et accessible, et souhaite fédérer une communauté d'utilisateurs et de développeurs autour de ces approches. L'objectif est de structurer des collaborations scientifiques et technologiques, notamment via des projets portant sur l'usage rigoureux des modèles spectro-chimiométriques en imagerie, en conditions passives de terrain ou actives en environnement industriel. Une explicitation du principe instrumental sera proposée, avec invitation à rejoindre cette dynamique et à contribuer au développement d'une nouvelle génération d'instruments hybrides, transformant le spectromètre en dispositif d'imagerie et de décision.

L'association HélioSPIR

remercie les sponsors qui contribuent à assurer le fonctionnement du réseau et le succès de ces rencontres annuelles

Sponsors Platinum

ARDOP INDUSTRIE

bonsai ADVANCED TECHNOLOGIES

BRUKER

BÜCHI SWITZERLAND

Fondis Electronic
Groupe Physitek

FOSS

IDIL FIBRES OPTIQUES

ondalys

PAT-INDUSTRY
UNDERSTAND. CONTROL. OPTIMIZE.

Polytec

PRO-LITE TECHNOLOGY

interline

Quantum Design

Sponsor Silver

PerkinElmer

Sponsors institutionnels

cirad

INRAE
la science pour la vie, l'humain, la terre

IRD
Institut de Recherche pour le Développement
FRANCE

PURPAN ÉCOLE D'INGÉNIEURS
Agriculture | Agroalimentaire
Sciences & Humanités